BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

Proc. Natl. Acad. Sci. USA
Vol. 95, pp. 6073–6078, May 1998
Biochemistry

Assessing sequence comparison methods with reliable structurally identified distant evolutionary relationships

STEVEN E. BRENNER*†‡, CYRUS CHOTHIA*, AND TIM J. P. HUBBARD§

*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, United Kingdom; and §Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambs CB10 1SA, United Kingdom

Communicated by David R. Davies, National Institute of Diabetes, Bethesda, MD, March 16, 1998 (received for review November 12, 1997)

ABSTRACT **Pairwise sequence comparison methods have been assessed using proteins whose relationships are known reliably from their structures and functions, as described in the SCOP database [Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia C. (1995) *J. Mol. Biol.* 247, 536–540]. The evaluation tested the programs BLAST [Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). *J. Mol. Biol.* 215, 403–410], WU-BLAST2 [Altschul, S. F. & Gish, W. (1996) *Methods Enzymol.* 266, 460–480], FASTA [Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* 85, 2444–2448], and SSEARCH [Smith, T. F. & Waterman, M. S. (1981) *J. Mol. Biol.* 147, 195–197] and their scoring schemes. The error rate of all algorithms is greatly reduced by using statistical scores to evaluate matches rather than percentage identity or raw scores. The E-value statistical scores of SSEARCH and FASTA are reliable: the number of false positives found in our tests agrees well with the scores reported. However, the P-values reported by BLAST and WU-BLAST2 exaggerate significance by orders of magnitude. SSEARCH, FASTA ktup = 1, and WU-BLAST2 perform best, and they are capable of detecting almost all relationships between proteins whose sequence identities are >30%. For more distantly related proteins, they do much less well; only one-half of the relationships between proteins with 20–30% identity are found. Because many homologs have low sequence similarity, most distant relationships cannot be detected by any pairwise comparison method; however, those which are identified may be used with confidence.**

Sequence database searching plays a role in virtually every branch of molecular biology and is crucial for interpreting the sequences issuing forth from genome projects. Given the method's central role, it is surprising that overall and relative capabilities of different procedures are largely unknown. It is difficult to verify algorithms on sample data because this requires large data sets of proteins whose evolutionary relationships are known unambiguously and independently of the methods being evaluated. However, nearly all known homologs have been identified by sequence analysis (the method to be tested). Also, it is generally very difficult to know, in the absence of structural data, whether two proteins that lack clear sequence similarity are unrelated. This has meant that although previous evaluations have helped improve sequence comparison, they have suffered from insufficient, imperfectly characterized, or artificial test data. Assessment also has been problematic because high quality database sequence searching attempts to have both sensitivity (detection of homologs) and specificity (rejection of unrelated proteins); however, these complementary goals are linked such that increasing one causes the other to be reduced.

Sequence comparison methodologies have evolved rapidly, so no previously published tests has evaluated modern versions of programs commonly used. For example, parameters in BLAST (1) have changed, and WU-BLAST2 (2)—which produces gapped alignments—has become available. The latest version of FASTA (3) previously tested was 1.6, but the current release (version 3.0) provides fundamentally different results in the form of statistical scoring.

The previous reports also have left gaps in our knowledge. For example, there has been no published assessment of thresholds for scoring schemes more sophisticated than percentage identity. Thus, the widely discussed statistical scoring measures have never actually been evaluated on large databases of real proteins. Moreover, the different scoring schemes commonly in use have not been compared.

Beyond these issues, there is a more fundamental question: in an absolute sense, how well does pairwise sequence comparison work? That is, what fraction of homologous proteins can be detected using modern database searching methods?

In this work, we attempt to answer these questions and to overcome both of the fundamental difficulties that have hindered assessment of sequence comparison methodologies. First, we use the set of distant evolutionary relationships in the SCOP: Structural Classification of Proteins database (4), which is derived from structural and functional characteristics (5). The SCOP database provides a uniquely reliable set of homologs, which are known independently of sequence comparison. Second, we use an assessment method that jointly measures both sensitivity and specificity. This method allows straightforward comparison of different sequence searching procedures. Further, it can be used to aid interpretation of real database searches and thus provide optimal and reliable results.

Previous Assessments of Sequence Comparison. Several previous studies have examined the relative performance of different sequence comparison methods. The most encompassing analyses have been by Pearson (6, 7), who compared the three most commonly used programs. Of these, the Smith–Waterman algorithm (8) implemented in SSEARCH (3) is the oldest and slowest but the most rigorous. Modern heuristics have provided BLAST (1) the speed and convenience to make it the most popular program. Intermediate between these two is FASTA (3), which may be run in two modes offering either greater speed (ktup = 2) or greater effectiveness (ktup = 1). Pearson also considered different parameters for each of these programs.

To test the methods, Pearson selected two representative proteins from each of 67 protein superfamilies defined by the PIR database (9). Each was used as a query to search the database, and the matched proteins were marked as being homologous or unrelated according to their membership of PIR

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: EPQ, errors per query.

†Present address: Department of Structural Biology, Stanford University, Fairchild Building D-109, Stanford, CA 94305-5126
‡To whom reprints requests should be addressed. e-mail: brenner@hyper.stanford.edu.

superfamilies. Pearson found that modern matrices and "ln-scaling" of raw scores improve results considerably. He also reported that the rigorous Smith–Waterman algorithm worked slightly better than FASTA, which was in turn more effective than BLAST.

Very large scale analyses of matrices have been performed (10), and Henikoff and Henikoff (11) also evaluated the effectiveness of BLAST and FASTA. Their test with BLAST considered the ability to detect homologs above a predetermined score but had no penalty for methods which also reported large numbers of spurious matches. The Henikoffs searched the SWISS-PROT database (12) and used PROSITE (13) to define homologous families. Their results showed that the BLOSUM62 matrix (14) performed markedly better than the extrapolated PAM-series matrices (15), which previously had been popular.

A crucial aspect of any assessment is the data that are used to test the ability of the program to find homologs. But in Pearson's and the Henikoffs' evaluations of sequence comparison, the correct results were effectively unknown. This is because the superfamilies in PIR and PROSITE are principally created by using the same sequence comparison methods which are being evaluated. Interdependency of data and methods creates a "chicken and egg" problem, and means for example, that new methods would be penalized for correctly identifying homologs missed by older programs. For instance, immunoglobulin variable and constant domains are clearly homologous, but PIR places them in different superfamilies. The problem is widespread: each superfamily in PIR 48.00 with a structural homolog is itself homologous to an average of 1.6 other PIR superfamilies (16).

To surmount these sorts of difficulties, Sander and Schneider (17) used protein structures to evaluate sequence comparison. Rather than comparing different sequence comparison algorithms, their work focused on determining a length-dependent threshold of percentage identity, above which all proteins would be of similar structure. A result of this analysis was the HSSP equation; it states that proteins with 25% identity over 80 residues will have similar structures, whereas shorter alignments require higher identity. (Other studies also have used structures (18–20), but these focused on a small number of model proteins and were principally oriented toward evaluating alignment accuracy rather than homology detection.)

A general solution to the problem of scoring comes from statistical measures (i.e., E-values and P-values) based on the extreme value distribution (21). Extreme value scoring was implemented analytically in the BLAST program using the Karlin and Altschul statistics (22, 23) and empirical approaches have been recently added to FASTA and SSEARCH. In addition to being heralded as a reliable means of recognizing significantly similar proteins (24, 25), the mathematical tractability of statistical scores "is a crucial feature of the BLAST algorithm" (1). The validity of this scoring procedure has been tested analytically and empirically (see ref. 2 and references in ref. 24). However, all large empirical tests used random sequences that may lack the subtle structure found within biological sequences (26, 27) and obviously do not contain any real homologs. Thus, although many researchers have suggested that statistical scores be used to rank matches (24, 25, 28), there have been no large rigorous experiments on biological data to determine the degree to which such rankings are superior.

A Database for Testing Homology Detection. Since the discovery that the structures of hemoglobin and myoglobin are very similar though their sequences are not (29), it has been apparent that comparing structures is a more powerful (if less convenient) way to recognize distant evolutionary relationships than comparing sequences. If two proteins show a high degree of similarity in their structural details and function, it is very probable that they have an evolutionary relationship though their sequence similarity may be low.

The recent growth of protein structure information combined with the comprehensive evolutionary classification in the SCOP database (4, 5) have allowed us to overcome previous limitations. With these data, we can evaluate the performance of sequence comparison methods on real protein sequences whose relationships are known confidently. The SCOP database uses structural information to recognize distant homologs, the large majority of which can be determined unambiguously. These superfamilies, such as the globins or the immunoglobulins, would be recognized as related by the vast majority of the biological community despite the lack of high sequence similarity.

From SCOP, we extracted the sequences of domains of proteins in the Protein Data Bank (PDB) (30) and created two databases. One (PDB90D-B) has domains, which were all <90% identical to any other, whereas (PDB40D-B) had those <40% identical. The databases were created by first sorting all protein domains in SCOP by their quality and making a list. The highest quality domain was selected for inclusion in the database and removed from the list. Also removed from the list (and discarded) were all other domains above the threshold level of identity to the selected domain. This process was repeated until the list was empty. The PDB40D-B database contains 1,323 domains, which have 9,044 ordered pairs of distant relationships, or ≈0.5% of the total 1,749,006 ordered pairs. In PDB90D-B, the 2,079 domains have 53,988 relationships, representing 1.2% of all pairs. Low complexity regions of sequence can achieve spurious high scores, so these were masked in both databases by processing with the SEG program (27) using recommended parameters: 12 1.8 2.0. The databases used in this paper are available from http://sss.stanford.edu/sss/, and databases derived from the current version of SCOP may be found at http://scop.mrc-lmb.cam.ac.uk/scop/.

Analyses from both databases were generally consistent, but PDB40D-B focuses on distantly related proteins and reduces the heavy overrepresentation in the PDB of a small number of families (31, 32), whereas PDB90D-B (with more sequences) improves evaluations of statistics. Except where noted otherwise, the distant homolog results here are from PDB40D-B. Although the precise numbers reported here are specific to the structural domain databases used, we expect the trends to be general.

Assessment Data and Procedure. Our assessment of sequence comparison may be divided into four different major categories of tests. First, using just a single sequence comparison algorithm at a time, we evaluated the effectiveness of different scoring schemes. Second, we assessed the reliability of scoring procedures, including an evaluation of the validity of statistical scoring. Third, we compared sequence comparison algorithms (using the optimal scoring scheme) to determine their relative performance. Fourth, we examined the distribution of homologs and considered the power of pairwise sequence comparison to recognize them. All of the analyses used the databases of structurally identified homologs and a new assessment criterion.

The analyses tested BLAST (1), version 1.4.9MP, and WU-BLAST2 (2), version 2.0a13MP. Also assessed was the FASTA package, version 3.0t76 (3), which provided FASTA and the SSEARCH implementation of Smith–Waterman (8). For SSEARCH and FASTA, we used BLOSUM45 with gap penalties −12/−1 (7, 16). The default parameters and matrix (BLOSUM62) were used for BLAST and WU-BLAST2.

The "Coverage Vs. Error" Plot. To test a particular protocol (comprising a program and scoring scheme), each sequence from the database was used as a query to search the database. This yielded ordered pairs of query and target sequences with associated scores, which were sorted, on the basis of their scores, from best to worst. The ideal method would have

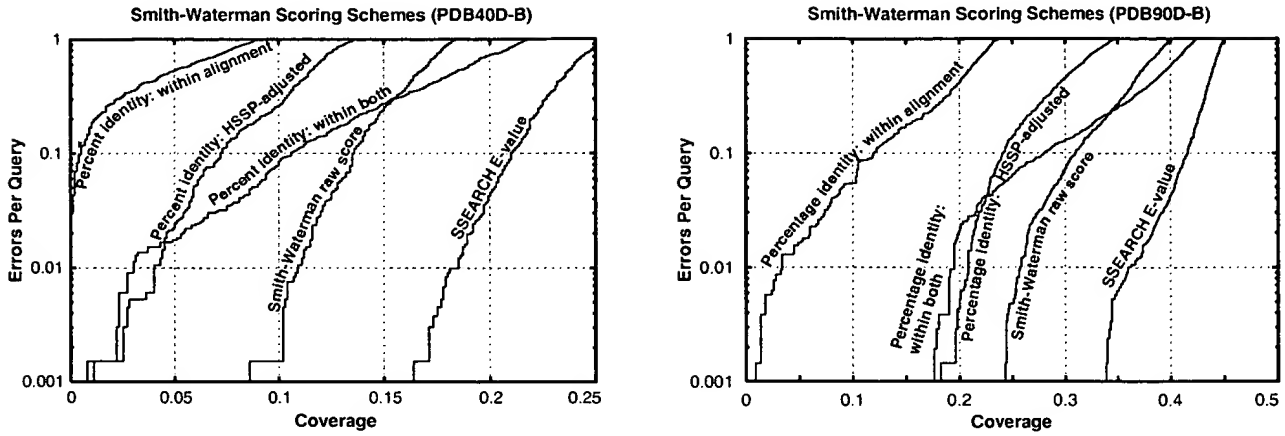


FIG. 1. Coverage vs. error plots of different scoring schemes for SSEARCH Smith–Waterman. (*A*) Analysis of PDB40D-B database. (*B*) Analysis of PDB90D-B database. All of the proteins in the database were compared with each other using the SSEARCH program. The results of this single set of comparisons were considered using five different scoring schemes and assessed. The graphs show the coverage and errors per query (EPQ) for statistical scores, raw scores, and three measures using percentage identity. In the coverage vs. error plot, the x axis indicates the fraction of all homologs in the database (known from structure) which have been detected. Precisely, it is the number of detected pairs of proteins with the same fold divided by the total number of pairs from a common superfamily. PDB40D-B contains a total of 9,044 homologs, so a score of 10% indicates identification of 904 relationships. The y axis reports the number of EPQ. Because there are 1,323 queries made in the PDB40D-B all-vs.-all comparison, 13 errors corresponds to 0.01, or 1% EPQ. The y axis is presented on a log scale to show results over the widely varying degrees of accuracy which may be desired. The scores that correspond to the levels of EPQ and coverage are shown in Fig. 4 and Table 1. The graph demonstrates the trade-off between sensitivity and selectivity. As more homologs are found (moving to the right), more errors are made (moving up). The ideal method would be in the lower right corner of the graph, which corresponds to identifying many evolutionary relationships without selecting unrelated proteins. Three measures of percentage identity are plotted. Percentage identity within alignment is the degree of identity within the aligned region of the proteins, without consideration of the alignment length. Percentage identity within both is the number of identical residues in the aligned region as a percentage of the average length of the query and target proteins. The HSSP equation (17) is $H = 290.15l^{-0.562}$ where l is length for $10 < l < 80$; $H > 100$ for $l < 10$; $H = 24.7$ for $l > 80$. The percentage identity HSSP-adjusted score is the percent identity within the alignment minus H. Smith–Waterman raw scores and E-values were taken directly from the sequence comparison program.

perfect separation, with all of the homologs at the top of the list and unrelated proteins below. In practice, perfect separation is impossible to achieve so instead one is interested in drawing a threshold above which there are the largest number of related pairs of sequences consistent with an acceptable error rate.

Our procedure involved measuring the coverage and error for every threshold. Coverage was defined as the fraction of structurally determined homologs that have scores above the selected threshold; this reflects the sensitivity of a method. Errors per query (EPQ), an indicator of selectivity, is the number of nonhomologous pairs above the threshold divided by the number of queries. Graphs of these data, called coverage vs. error plots, were devised to understand how protocols compare at different levels of accuracy. These graphs share effectively all of the beneficial features of Reciever Operating Characteristic (ROC) plots (33, 34) but better represent the high degrees of accuracy required in sequence comparison and the huge background of nonhomologs.

This assessment procedure is directly relevant to practical sequence database searching, for it provides precisely the information necessary to perform a reliable sequence database search. The EPQ measure places a premium on score consistency; that is, it requires scores to be comparable for different queries. Consistency is an aspect which has been largely

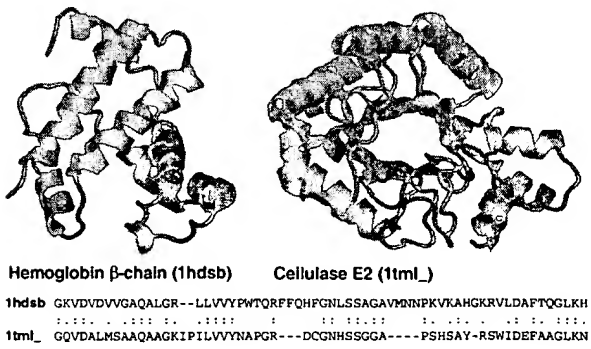


FIG. 2. Unrelated proteins with high percentage identity. Hemoglobin β-chain (PDB code 1hds chain b, ref. 38, *Left*) and cellulase E2 (PDB code 1tml, ref. 39, *Right*) have 39% identity over 64 residues, a level which is often believed to be indicative of homology. Despite this high degree of identity, their structures strongly suggest that these proteins are not related. Appropriately, neither the raw alignment score of 85 nor the E-value of 1.3 is significant. Proteins rendered by RASMOL (40).

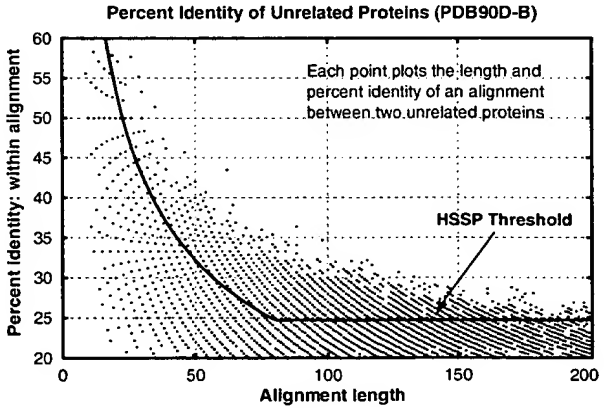


FIG. 3. Length and percentage identity of alignments of unrelated proteins in PDB90D-B: Each pair of nonhomologous proteins found with SSEARCH is plotted as a point whose position indicates the length and the percentage identity within the alignment. Because alignment length and percentage identity are quantized, many pairs of proteins may have exactly the same alignment length and percentage identity. The line shows the HSSP threshold (though it is intended to be applied with a different matrix and parameters).

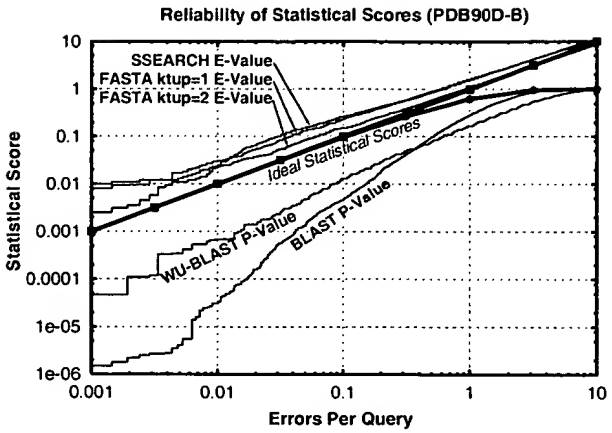


FIG. 4. Reliability of statistical scores in PDB90D-B: Each line shows the relationship between reported statistical score and actual error rate for a different program. E-values are reported for SSEARCH and FASTA, whereas P-values are shown for BLAST and WU-BLAST2. If the scoring were perfect, then the number of errors per query and the E-values would be the same, as indicated by the upper bold line. (P-values should be the same as EPQ for small numbers, and diverges at higher values, as indicated by the lower bold line.) E-values from SSEARCH and FASTA are shown to have good agreement with EPQ but underestimate the significance slightly. BLAST and WU-BLAST2 are overconfident, with the degree of exaggeration dependent upon the score. The results for PDB40D-B were similar to those for PDB90D-B despite the difference in number of homologs detected. This graph could be used to roughly calibrate the reliability of a given statistical score.

ignored in previous tests but is essential for the straightforward or automatic interpretation of sequence comparison results. Further, it provides a clear indication of the confidence that should be ascribed to each match. Indeed, the EPQ measure should approximate the expectation value reported by database searching programs, if the programs' estimates are accurate.

The Performance of Scoring Schemes. All of the programs tested could provide three fundamental types of scores. The first score is the percentage identity, which may be computed in several ways based on either the length of the alignment or the lengths of the sequences. The second is a "raw" or "Smith–Waterman" score, which is the measure optimized by the Smith–Waterman algorithm and is computed by summing the substitution matrix scores for each position in the alignment and subtracting gap penalties. In BLAST, a measure related to this score is scaled into bits. Third is a statistical score based on the extreme value distribution. These results are summarized in Fig. 1.

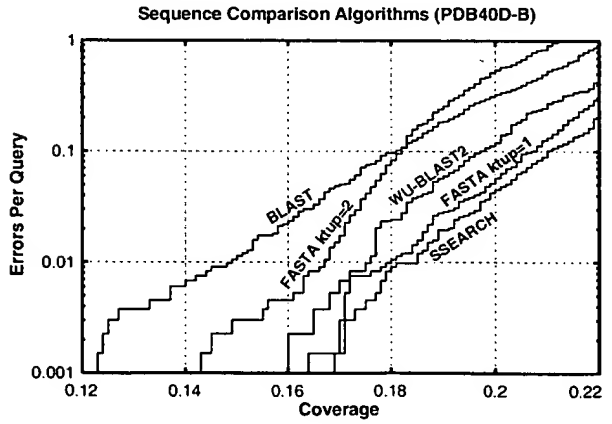
Sequence Identity. Though it has been long established that percentage identity is a poor measure (35), there is a common rule-of-thumb stating that 30% identity signifies homology. Moreover, publications have indicated that 25% identity can be used as a threshold (17, 36). We find that these thresholds, originally derived years ago, are not supported by present results. As databases have grown, so have the possibilities for chance alignments with high identity; thus, the reported cutoffs lead to frequent errors. Fig. 2 shows one of the many pairs of proteins with very different structures that nonetheless have high levels of identity over considerable aligned regions. Despite the high identity, the raw and the statistical scores for such incorrect matches are typically not significant. The principal reasons percentage identity does so poorly seem to be that it ignores information about gaps and about the conservative or radical nature of residue substitutions.

From the PDB90D-B analysis in Fig. 3, we learn that 30% identity is a reliable threshold for this database only for sequence alignments of at least 150 residues. Because one unrelated pair of proteins has 43.5% identity over 62 residues, it is probably necessary for alignments to be at least 70 residues in length before 40% is a reasonable threshold, for a database of this particular size and composition.

At a given reliability, scores based on percentage identity detect just a fraction of the distant homologs found by statistical scoring. If one measures the percentage identity in the aligned regions without consideration of alignment length, then a negligible number of distant homologs are detected. Use of the HSSP equation improves the value of percentage identity, but even this measure can find only 4% of all known homologs at 1% EPQ. In short, percentage identity discards most of the information measured in a sequence comparison.

Raw Scores. Smith–Waterman raw scores perform better than percentage identity (Fig. 1), but ln-scaling (7) provided no notable benefit in our analysis. It is necessary to be very precise when using either raw or bit scores because a 20% change in cutoff score could yield a tenfold difference in EPQ. However, it is difficult to choose appropriate thresholds because the reliability of a bit score depends on the lengths of the proteins matched and the size of the database. Raw score thresholds also are affected by matrix and gap parameters.

Statistical Scores. Statistical scores were introduced partly to overcome the problems that arise from raw scores. This scoring scheme provides the best discrimination between homologous proteins and those which are unrelated. Most



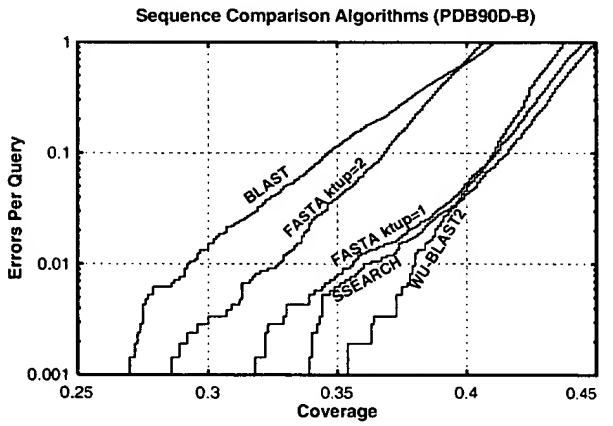


FIG. 5. Coverage vs. error plots of different sequence comparison methods: Five different sequence comparison methods are evaluated, each using statistical scores (E- or P-values). (*A*) PDB40D-B database. In this analysis, the best method is the slow SSEARCH, which finds 18% of relationships at 1% EPQ. FASTA ktup = 1 and WU-BLAST2 are almost as good. (*B*) PDB90D-B database. The quick WU-BLAST2 program provides the best coverage at 1% EPQ on this database, although at higher levels of error it becomes slightly worse than FASTA ktup = 1 and SSEARCH.

likely, its power can be attributed to its incorporation of more information than any other measure; it takes account of the full substitution and gap data (like raw scores) but also has details about the sequence lengths and composition and is scaled appropriately.

We find that statistical scores are not only powerful, but also easy to interpret. SSEARCH and FASTA show close agreement between statistical scores and actual number of errors per query (Fig. 4). The expectation value score gives a good, slightly conservative estimate of the chances of the two sequences being found at random in a given query. Thus, an E-value of 0.01 indicates that roughly one pair of nonhomologs of this similarity should be found in every 100 different queries. Neither raw scores nor percentage identity can be interpreted in this way, and these results validate the suitability of the extreme value distribution for describing the scores from a database search.

The P-values from BLAST also should be directly interpretable but were found to overstate significance by more than two orders of magnitude for 1% EPQ for this database. Nonetheless, these results strongly suggest that the analytic theory is fundamentally appropriate. WU-BLAST2 scores were more reliable than those from BLAST, but also exaggerate expected confidence by more than an order of magnitude at 1% EPQ.

Overall Detection of Homologs and Comparison of Algorithms. The results in Fig. 5*A* and Table 1 show that pairwise sequence comparison is capable of identifying only a small fraction of the homologous pairs of sequences in PDB40D-B. Even SSEARCH with E-values, the best protocol tested, could find only 18% of all relationships at a 1% EPQ. BLAST, which identifies 15%, was the worst performer, whereas FASTA ktup = 1 is nearly as effective as SSEARCH. FASTA ktup = 2 and WU-BLAST2 are intermediate in their ability to detect homologs. Comparison of different algorithms indicates that those capable of identifying more homologs are generally slower. SSEARCH is 25 times slower than BLAST and 6.5 times slower than FASTA ktup = 1. WU-BLAST2 is slightly faster than FASTA ktup = 2, but the latter has more interpretable scores.

In PDB90D-B, where there are many close relationships, the best method can identify only 38% of structurally known homologs (Fig. 5*B*). The method which finds that many relationships is WU-BLAST2. Consequently, we infer that the differences between FASTA kup = 1, SSEARCH, and WU-BLAST2 programs are unlikely to be significant when compared with variation in database composition and scoring reliability.

Fig. 6 helps to explain why most distant homologs cannot be found by sequence comparison: a great many such relationships have no more sequence identity than would be expected by chance. SSEARCH with E-values can recognize >90% of the homologous pairs with 30–40% identity. In this region, there are 30 pairs of homologous proteins that do not have significant E-values, but 26 of these involve sequences with <50 residues. Of sequences having 25–30% identity, 75% are identified by SSEARCH E-values. However, although the number of homologs grows at lower levels of identity, the detection falls off sharply: only 40% of homologs with 20–25% identity

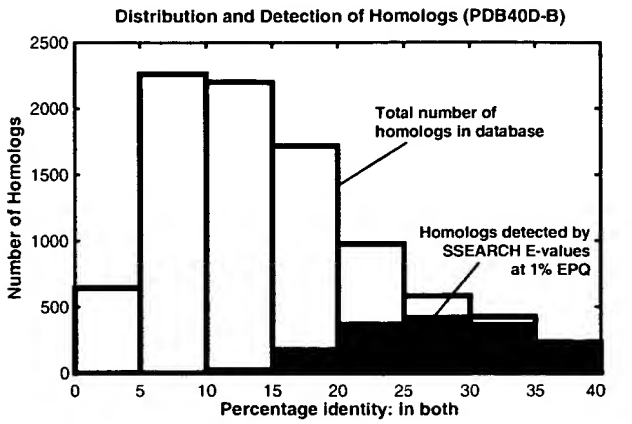


FIG. 6. Distribution and detection of homologs in PDB40D-B. Bars show the distribution of homologous pairs PDB40D-B according to their identity (using the measure of identity in both). Filled regions indicate the number of these pairs found by the best database searching method (SSEARCH with E-values) at 1% EPQ. The PDB40D-B database contains proteins with <40% identity, and as shown on this graph, most structurally identified homologs in the database have diverged extremely far in sequence and have <20% identity. Note that the alignments may be inaccurate, especially at low levels of identity. Filled regions show that SSEARCH can identify most relationships that have 25% or more identity, but its detection wanes sharply below 25%. Consequently, the great sequence divergence of most structurally identified evolutionary relationships effectively defeats the ability of pariwise sequence comparison to detect them.

are detected and only 10% of those with 15–20% can be found. These results show that statistical scores can find related proteins whose identity is remarkably low; however, the power of the method is restricted by the great divergence of many protein sequences.

After completion of this work, a new version of pairwise BLAST was released: BLASTGP (37). It supports gapped alignments, like WU-BLAST2, and dispenses with sum statistics. Our initial tests on BLASTGP using default parameters show that its E-values are reliable and that its overall detection of homologs was substantially better than that of ungapped BLAST, but not quite equal to that of WU-BLAST2.

CONCLUSION

The general consensus amongst experts (see refs. 7, 24, 25, 27 and references therein) suggests that the most effective sequence searches are made by (*i*) using a large current database in which the protein sequences have been complexity masked and (*ii*) using statistical scores to interpret the results. Our experiments fully support this view.

Our results also suggest two further points. First, the E-values reported by FASTA and SSEARCH give fairly accurate estimates of the significance of each match, but the P-values provided by BLAST and WU-BLAST2 underestimate the true

Table 1. Summary of sequence comparison methods with PDB40D-B

Method	Relative Time*	1% EPQ Cutoff	Coverage at 1% EPQ
SSEARCH % identity: within alignment	25.5	>70%	<0.1
SSEARCH % identity: within both	25.5	34%	3.0
SSEARCH % identity: HSSP-scaled	25.5	35% (HSSP + 9.8)	4.0
SSEARCH Smith–Waterman raw scores	25.5	142	10.5
SSEARCH E-values	25.5	0.03	18.4
FASTA ktup = 1 E-values	3.9	0.03	17.9
FASTA ktup = 2 E-values	1.4	0.03	16.7
WU-BLAST2 P-values	1.1	0.003	17.5
BLAST P-values	1.0	0.00016	14.8

*Times are from large database searches with genome proteins.

extent of errors. Second, SSEARCH, WU-BLAST2, and FASTA ktup = 1 perform best, though BLAST and FASTA ktup = 2 detect most of the relationships found by the best procedures and are appropriate for rapid initial searches.

The homologous proteins that are found by sequence comparison can be distinguished with high reliability from the huge number of unrelated pairs. However, even the best database searching procedures tested fail to find the large majority of distant evolutionary relationships at an acceptable error rate. Thus, if the procedures assessed here fail to find a reliable match, it does not imply that the sequence is unique; rather, it indicates that any relatives it might have are distant ones.**

**Additional and updated information about this work, including supplementary figures, may be found at http://sss.stanford.edu/sss/.

The authors are grateful to Drs. A. G. Murzin, M. Levitt, S. R. Eddy, and G. Mitchison for valuable discussion. S.E.B. was principally supported by a St. John's College (Cambridge, UK) Benefactors' Scholarship and by the American Friends of Cambridge University. S.E.B. dedicates his contribution to the memory of Rabbi Albert T. and Clara S. Bilgray.

1. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* **215,** 403–410.
2. Altschul, S. F. & Gish, W. (1996) *Methods Enzymol.* **266,** 460–480.
3. Pearson, W. R. & Lipman, D. J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2444–2448.
4. Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995) *J. Mol. Biol.* **247,** 536–540.
5. Brenner, S. E., Chothia, C., Hubbard, T. J. P. & Murzin, A. G. (1996) *Methods Enzymol.* **266,** 635–643.
6. Pearson, W. R. (1991) *Genomics* **11,** 635–650.
7. Pearson, W. R. (1995) *Protein Sci.* **4,** 1145–1160.
8. Smith, T. F. & Waterman, M. S. (1981) *J. Mol. Biol.* **147,** 195–197.
9. George, D. G., Hunt, L. T. & Barker, W. C. (1996) *Methods Enzymol.* **266,** 41–59.
10. Vogt, G., Etzold, T. & Argos, P. (1995) *J. Mol. Biol.* **249,** 816–831.
11. Henikoff, S. & Henikoff, J. G. (1993) *Proteins* **17,** 49–61.
12. Bairoch, A. & Apweiler, R. (1996) *Nucleic Acids Res.* **24,** 21–25.
13. Bairoch, A., Bucher, P. & Hofmann, K. (1996) *Nucleic Acids Res.* **24,** 189–196.
14. Henikoff, S. & Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10915–10919.
15. Dayhoff, M., Schwartz, R. M. & Orcutt, B. C. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. (National Biomedical Research Foundation, Silver Spring, MD), Vol. 5, Suppl. 3, pp. 345–352.
16. Brenner, S. E. (1996) Ph.D. thesis. (University of Cambridge, UK).
17. Sander, C. & Schneider, R. (1991) *Proteins* **9,** 56–68.
18. Johnson, M. S. & Overington, J. P. (1993) *J. Mol. Biol.* **233,** 716–738.
19. Barton, G. J. & Sternberg, M. J. E. (1987) *Protein Eng.* **1,** 89–94.
20. Lesk, A. M., Levitt, M. & Chothia, C. (1986) *Protein Eng.* **1,** 77–78.
21. Arratia, R., Gordon, L. & M, W. (1986) *Ann. Stat.* **14,** 971–993.
22. Karlin, S. & Altschul, S. F. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 2264–2268.
23. Karlin, S. & Altschul, S. F. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 5873–5877.
24. Altschul, S. F., Boguski, M. S., Gish, W. & Wootton, J. C. (1994) *Nat. Genet.* **6,** 119–129.
25. Pearson, W. R. (1996) *Methods Enzymol.* **266,** 227–258.
26. Lipman, D. J., Wilbur, W. J., Smith, T. F. & Waterman, M. S. (1984) *Nucleic Acids Res.* **12,** 215–226.
27. Wootton, J. C. & Federhen, S. (1996) *Methods Enzymol.* **266,** 554–571.
28. Waterman, M. S. & Vingron, M. (1994) *Stat. Science* **9,** 367–381.
29. Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) *J. Mol. Biol.* **13,** 669–678.
30. Abola, E. E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987) in *Crystallographic Databases: Information Content, Software Systems, Scientific Applications*, eds. Allen, F. H., Bergerhoff, G. & Sievers, R. (Data Comm. Intl. Union Crystallogr., Cambridge, UK), pp. 107–132.
31. Brenner, S. E., Chothia, C. & Hubbard, T. J. P. (1997) *Curr. Opin. Struct. Biol.* **7,** 369–376.
32. Orengo, C., Michie, A., Jones S, Jones D. T, Swindells M. B. & Thornton, J. (1997) *Structure (London)* **5,** 1093–1108.
33. Zweig, M. H. & Campbell, G. (1993) *Clin. Chem.* **39,** 561–577.
34. Gribskov, M. & Robinson, N. L. (1996) *Comput. Chem.* **20,** 25–33.
35. Fitch, W. M. (1966) *J. Mol. Biol.* **16,** 9–16.
36. Chung, S. Y. & Subbiah, S. (1996) *Structure (London)* **4,** 1123–1127.
37. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
38. Girling, R., Schmidt, W., Jr, Houston, T., Amma, E. & Huisman, T. (1979) *J. Mol. Biol.* **131,** 417–433.
39. Spezio, M., Wilson, D. & Karplus, P. (1993) *Biochemistry* **32,** 9906–9916
40. Sayle, R. A. & Milner-White, E. J. (1995) *Trends Biochem. Sci.* **20,** 374–376.

0090-9556/04/3203-281–290$20.00
DRUG METABOLISM AND DISPOSITION

DMD 32:281–290, 2004

Vol. 32, No. 3
1299/1135075
Printed in U.S.A.

Symposium Article

GLUCURONIDATION AND THE UDP-GLUCURONOSYLTRANSFERASES IN HEALTH AND DISEASE

Peter G. Wells,[1] Peter I. Mackenzie,[1] Jayanta Roy Chowdhury,[1] Chantal Guillemette,[1] Philip A. Gregory, Yuji Ishii, Antony J. Hansen, Fay K. Kessler, Perry M. Kim, Namita Roy Chowdhury, and Joseph K. Ritter[1]

Faculty of Pharmacy and Department of Pharmacology, University of Toronto, Toronto, Ontario, Canada (P.G.W., P.M.K.); Department of Clinical Pharmacology, Flinders Medical Centre, Adelaide, South Australia, Australia (P.I.M., P.A.G., A.J.H.); Graduate School of Pharmaceutical Sciences, Kyushu University (Y.I.), Fukuoka-shi, Kyushu, Japan; Albert Einstein College of Medicine, Yeshiva University, Bronx, New York (J.R.C., N.R.C.); Faculty of Pharmacy, Laval University, Quebec City, Quebec, Canada (C.G.); and Department of Pharmacology and Toxicology, Virginia Commonwealth University, Richmond, Virginia (J.K.R., F.K.K.)

(Received October 10, 2003; accepted December 11, 2003)

This article is available online at http://dmd.aspetjournals.org

ABSTRACT:

This article is an updated report of a symposium held at the June 2000 annual meeting of the American Society for Pharmacology and Experimental Therapeutics in Boston. The symposium was sponsored by the ASPET Divisions for Drug Metabolism and Molecular Pharmacology. The report covers research from the authors' laboratories on the structure and regulation of UDP-glucuronosyltransferase (UGT) genes, glucuronidation of xenobiotics and endobiotics, the toxicological relevance of UGTs, the role of UGT polymorphisms in cancer susceptibility, and gene therapy for UGT deficiencies.

For most xenobiotics and many endobiotics, glucuronidation constitutes a major route of elimination and thereby may substantially modulate substrate concentrations and effects. In some cases, glucuronidation forms the biologically active molecule. Recent studies have revealed an extensive superfamily of UDP-glucuronosyltransferases (UGTs),[2] previously termed glucuronyl transferases, which catalyze the conjugation of UDP-glucuronic acid with lipid-soluble substrates to form polar conjugates that are excreted in the urine and feces. These studies have provided fundamental insights into UGT gene structure and regulation, isozyme substrate selectivity, and interindividual variability. Whereas there remains much to learn about the potential biological relevance of UGTs, deficient glucuronidation can result in either elevated tissue concentrations and direct toxicity of substrates, as with the endobiotic bilirubin, or, alternatively, enhanced bioactivation of substrates to toxic reactive intermediates, as in the case of acetaminophen and benzo[*a*]pyrene. Interindividual UGT variability likely plays an important role in drug efficacy and xenobiotic toxicity, as well as in hormonal regulation and certain diseases, which in some cases may be amenable to therapeutic manipulations including gene therapy.

This work was supported in part by grants from the Canadian Institutes of Health Research (P.G.W., C.G.), the National Health and Medical Research Council of Australia and the Anti-Cancer Foundation of South Australia (P.I.M., P.A.G., A.J.H., Y.I.), the Canada Research Chair Program and the Fonds de la Recherche en Sante du Quebec (C.G.), and United States Public Health Service Grants R01-DK46057, R01-DK34357 and P30DK41296 (J.R.C., N.R.C.), and R01ES07762 (J.K.R.).

[1] These authors contributed equally as symposium speakers.

[2] Abbreviations used are: UGT, UDP-glucuronosyltransferase; B[a]P, benzo[a]pyrene; CN-1, Crigler-Najjar syndrome type 1; HCA, heterocyclic amine; HNF1, hepatocyte nuclear factor 1; OR, odds ratio; SN-38, 7-ethyl-10-hydroxycamptothecin; NNK, 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone; P450, cytochrome P450; rAAV, recombinant adenoassociated virus.

Address correspondence to: Dr. Joseph K. Ritter, Dept. of Pharmacology and Toxicology, Medical College of Virginia, Virginia Commonwealth University, 410 N. 12th St., P.O. Box 980613, Richmond, VA 23298-0613. E-mail: jritter@mail2.vcu.edu

Structure and Tissue-Specific Regulation of UGT Genes (P.I.M., P.A.G., Y.I., A.J.H.)

The UGT content of cells and tissues is a major determinant of our response to those chemicals that are primarily eliminated by conjugation with glucuronic acid. There are marked interindividual differences in the content of UGTs in the liver and other organs including the gastrointestinal tract. For example, only one third of the population appears to express UGT1A1, UGT1A3, and UGT1A6 in their gastric epithelium (Strassburg et al., 1998). Studies on the mechanisms that regulate *UGT* genes, in a temporal and tissue-specific manner, should contribute significantly to understanding the basis for these differences. Such studies should also aid in the design of molecular probes to assess the capacity of individuals to metabolize specific drugs and toxins, before exposure to these agents.

The genes encoding UGTs that use UDP-glucuronic acid as sugar donor have been assigned to two families (Mackenzie et al., 1997). The *UGT1* family constitutes a complex gene locus on human chromosome 2q37 and comprises 13 first exons that encode the unique

N-terminal domains of the UGT1A proteins and exons 2 to 5 that encode the C-terminal domain, which is identical in all UGT1A family members (Owens and Ritter, 1992). The UGT2 family, in contrast, is encoded by separate genes clustered on chromosome 4q13 and consists of the UGT2A and UGT2B subfamilies. The *UGT2B* genes that have been analyzed to date, viz., rat UGT2B1 (Mackenzie and Rodbourn 1990) and UGT2B2 (Haque et al., 1991), and human UGT2B4 (Monaghan et al., 1997), UGT2B7 (Ishii et al., 2000), UGT2B15 (Turgeon et al., 2000), and UGT2B17 (Beaulieu et al., 1997), consist of six exons and have highly similar exon/intron boundaries. Accumulating evidence demonstrates that many of the 17 human UGTs characterized to date exhibit tissue-specific patterns of expression. For example, UGT1A3 is found in the liver, kidney, and prostate, and throughout the gastrointestinal tract, whereas UGT1A8 is mainly restricted to the colon (Mojarrabi and Mackenzie, 1998). The factors that govern this specificity of UGT expression remain largely unknown.

Studies on rodent *UGT* genes have demonstrated that the transcription factors hepatocyte nuclear factor 1 (HNF1) and CAAT-enhancer binding protein are important positive regulators of *UGT* expression in the liver (Hansen et al., 1997, 1998). These nuclear proteins, which are members of the homeodomain and basic region-leucine zipper groups of transcription factors, respectively, have been shown to be important in the liver-specific expression of other genes such as albumin (Lichtsteiner et al., 1987). Furthermore, mice in which the hepatic expression of these factors has been substantially reduced display deficiencies in hepatic UGT1A1 and UGT2B expression (Lee et al., 1997). With the recent isolation and characterization of several human *UGT* genes, evidence is accumulating that many of the factors that contribute to liver-specific expression of UGTs in rodents are also important in regulating UGTs in human liver. This is exemplified by studies on the human *UGT2B7* gene (Ishii et al., 2000). Studies on this *UGT* gene were initiated since UGT2B7 glucuronidates many important therapeutic drugs, including nonsteroidal anti-inflammatory drugs, morphine, oxazepam, and zidovudine, and is found in both the liver and the gastrointestinal tract (Radominska-Pandya et al., 1999).

The proximal promoter of the *UGT2B7* gene contains an A/T-rich region just upstream from the transcription start site. Cotransfection experiments in the human liver-derived cell line, HepG2, and gel electrophoretic mobility shift assays demonstrated that HNF1α bound to this region to enhance *UGT2B7* promoter activity. The closely related transcription factor HNF1β had little effect on *UGT2B7* promoter activity. Furthermore, mutations in the A/T-rich region abolished the capacity of HNF1α to enhance promoter activity. Further investigation is required to define other factors that regulate the *UGT2B7* gene, either independently or via interactions with HNF1α. In this respect, it is interesting to note that the ubiquitous transcription factor, octamer transcription factor-1, acts synergistically with HNF1α to enhance *UGT2B7* promoter activity in HepG2 cells. Octamer transcription factor-1 is another member of the homeodomain group of transcription factors. It acts as a positive regulator of UGT2B7 expression by tethering to HNF1α and presumably promoting assembly of the RNA polymerase complex, rather than by directly binding to DNA (Ishii et al., 2000).

A comparison of the proximal promoter sequences of several human *UGT* genes reveals the presence of a similar A/T-rich region in many of these genes (Fig. 1). The binding of HNF1α to this sequence has been confirmed for the *UGT1A1* and *UGT2B17* genes (Bernard et al., 1999; Gregory et al., 2000). It is interesting to note that this putative HNF1 binding site is present in a similar position in the proximal promoters of those *UGT* genes that are expressed in the liver, but it is not present in the same position in those genes expressed exclusively in extrahepatic tissues.

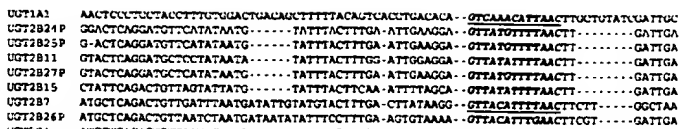


FIG. 1. *Alignment of UGT proximal promoter sequences.*

The sequences, including those of UGT pseudogenes (denoted by P), were obtained from GenBank, accession numbers AF180372 (UGT1A1), M84126 (UGT1A2P), M84127 (UGT1A3), M84128 (UGT1A4), M84129 (UGT1A5), AF014112 (UGT1A6), U39570 (UGT1A7), U42604 (UGT1A8), U39550 (UGT1A10), U39551 (UGT1A11P), U39552 (UGT1A12P), AF179875 (UGT2B4), AF282881 (UGT2B7), AF179873 (UGT2B11), AF179881 (UGT2B15), AF179874 (UGT2B17), AF179876 (UGT2B24P), AF179787 (UGT2B25P), AF179878 (UGT2B26P), AF179879 (UGT2B27P), and AF179880 (UGT2B28P). The putative HNF1 binding site is shown in italics. The sites that have been demonstrated experimentally to bind HNF1 protein are underlined. Note that UGT1A7, UGT1A8, and UGT1A10, which are not expressed in the liver, do not contain a consensus HNF1 binding site in a position equivalent to that of the other genes.

Although many UGTs are expressed in extrahepatic tissues, the mechanisms that regulate their tissue-specific expression have not been identified. The prostate contains UGT2B17, which is thought to play an important role in regulating intracellular androgen concentrations (Guillemette et al., 1997). The prostate-derived LNCap cell line contains large amounts of this enzyme and, hence, is a suitable model to investigate UGT2B17 regulation. The UGT2B17 gene promoter is substantially activated by exogenous HNF1α in transfected LNCaP cells. However, unlike HepG2 cells, LNCaP cells do not contain endogenous HNF1α but do contain the related transcription factor HNF1β. The latter appears to be a positive regulator of UGT2B17 expression in these cells and exerts its effects through a site that is distal to the HNF1α binding site utilized in HepG2 cells (Gregory et al., 2000). Factors specifying the expression of UGTs in other tissues remain to be identified.

Much work is required to define the mechanisms that contribute to the differential expression of UGT genes in the liver and other organs. Nevertheless, it is now evident that different factors and different combinations of factors regulate the same promoter in different cell types.

Glucuronidation of Xenobiotics and Endobiotics (J.K.R., F.K.K.)

Progress is continuing in cloning and establishing the substrate specificities of the individual UGT enzymes that catalyze this reaction. Their identification is essential for understanding why individuals vary in their rates of glucuronidation, altering the pharmacological and toxicological responses to different agents. Currently, there are ~20 distinct UGTs known in both humans and rats, the best characterized animal species. These can be divided into two families, UGT1 and UGT2, on the basis of sequence homology. Studies in our laboratory have focused on the substrate specificity and regulation of UGT1 isoforms.

The UGT1A1 form appears to be the most abundant UGT1 isoform expressed in liver (Ritter et al., 1992; Ikushiro et al., 1995) and is the principal isoform involved in the glucuronidation of bilirubin (Bosma et al., 1994). UGT1A1 is also selectively active toward certain phe-

nols [e.g., SN-38 (Iyer et al., 1998)] and 17α-ethinylestradiol (Ebner et al., 1993; Senafi et al., 1994). The available evidence points to high interindividual variability in UGT1A1 expression. In a study of UGT1A1 variation in human liver donor samples, marked variation (>50-fold) was observed in the microsomal UGT1A1 protein level (Ritter et al., 1999). Three samples with the highest UGT1A1 level were from patients who had received the anticonvulsant inducing agent, phenytoin, during their hospitalization. These findings are consistent with a report of elevated UGT1A1 mRNA in the liver of a phenytoin-exposed patient. They also agree with clinical evidence supporting the effectiveness of phenobarbital in inducing bilirubin elimination in patients with unconjugated hyperbilirubinemia.

Further evidence in support of environmental influences on UGT1A1 was provided by studies using primary human hepatocytes. After adaptation to culture, a common environment, UGT1A1 mRNA levels showed lower variability (<3- versus 50-fold). The three cultures from the phenytoin-exposed donors showed the most pronounced declines in UGT1A1 mRNA, likely due to the removal of inducing stimuli (phenytoin, diet). Induction studies showed that phenobarbital and oltipraz (prototypical inducing agents) resulted in elevated UGT1A1 mRNA, but exposure to 3-methylcholanthrene resulted in the most potent inducing effects (3- to 6-fold). These findings suggest that expression of UGT1A1 in human liver is under the control of multiple control mechanisms including the aryl hydrocarbon receptor. Exposure to polycyclic hydrocarbon-type inducers via cigarette smoking is another potential cause of UGT1A1 variation.

The possible role of the UGT1A1 promoter polymorphism (Bosma et al., 1995) in the observed variation was also investigated by genotyping the donors for the number of TA repeats in their UGT1A1 TATA boxes. A genetic influence on expression was supported by the observation that two individuals with the lowest UGT1A1 expression were homozygotes for the $(TA)_7TAA$ allele (Ritter et al., 1999). A third $(TA)_7TAA$ homozygote was one of the phenytoin-exposed patients. This patient exhibited lower UGT1A1 levels than did the two other phenytoin-exposed patients, who were both homozygous for the wild-type allele [$(TA)_6TAA$]. These data provide support for both genetic and environmental factors in interindividual variation in hepatic UGT1A1 expression.

The Gunn rat provides a useful animal model to investigate the relative contribution of UGT1 isoforms in total glucuronidating activities. The frame-shift mutation associated with the loss of UGT1A1 activity and hyperbilirubinemia in Gunn rats also inactivates the other UGT1 family isoforms. The contributions of UGT1 isozymes were assessed using various in vitro (microsomal UGT assays) and in vivo approaches (pharmacokinetics and organ toxicity). Two examples are the analgesic and potential hepatotoxicant, acetaminophen (de Morais et al., 1992a), and the toxic environmental pollutant, benzo[*a*]pyrene (B[a]P) (Hu and Wells, 1992). Establishing the identities of UGT1 isoforms involved in the glucuronidation of specific substrates requires the use of cloned and expressed UGT cDNAs. Using human embryonic kidney cells expressing the major UGT1 isoforms found in rat liver [UGT1A1, UGT1A5, UGT1A6, and UGT1A7 (Ikushiro et al., 1995)], we investigated the selectivities of these isoforms in the glucuronidation of bilirubin, acetaminophen, and B[a]P metabolites. Only the UGT1A1 isozyme was active toward bilirubin. In contrast, two of the four rat liver isoforms tested were active toward acetaminophen (UGT1A6 and UGT1A7) (Kessler et al., 2002). In rats maintained on a standard laboratory diet, UGT1A6 and UGT1A7 are expressed at low levels in liver but are induced after exposure to certain inducers (Grove et al., 1997; Kessler and Ritter, 1997; Kobayashi et al., 1998). These data likely indicate an important role for UGT1A6 and UGT1A7 in protection against acetaminophen-induced hepatotoxicity. The activities of these isoforms resembles the reported activities of human UGT1A6 and UGT1A9 toward acetaminophen (Bock et al., 1993). Apparent differences in the affinity and/or capacity of UGT1A6 (high affinity, low capacity) and UGT1A9 (low affinity, high capacity) for acetaminophen likely indicate that they will contribute differently to protection against acetaminophen in overdose situations. However, the possibility that other UGT1 isoforms besides the phenol UGTs contribute to acetaminophen glucuronidation is suggested by the finding that human UGT1A1 is also significantly active (Court et al., 2001). These results highlight the species-specific nature of drug glucuronidation.

Glucuronidation also is known to modulate toxicities associated with B[a]P exposure (Hu and Wells, 1992; Hu and Wells, 1994]. The high activity of the rat UGT1A7 form toward B[a]P metabolites, including phenols, quinols, and dihydrodiols, has been demonstrated (Grove et al., 1997). Despite its inducibility by polycyclic hydrocarbons, UGT1A6 shows very low activity toward most B[a]P metabolites, a finding consistent with the preference of UGT1A6 for small phenolic compounds with simple ring substitutions. Interestingly, rat UGT1A1 was found to be significantly active. Qualitatively, UGT1A1 resembles the UGT1A7 form in its activity toward many different B[a]P metabolites including B[a]P-7,8-dihydrodiol. These findings are supported by the observation that the human bilirubin UGT is active toward B[a]P-7,8-dihydrodiol(Fang et al., 2002). Although the specific activities of UGT1A1 were generally 2- to 6-fold lower than that of UGT1A7, some activities were higher (e.g., toward B[a]P-4,5-dihydrodiol). The high natural abundance of this form in liver, together with the observation of UGT1A1 inducibility by polycyclic aromatic hydrocarbons, supports a role for UGT1A1 deficiency in the mechanism of increased sensitivity of Gunn rats to B[a]P-induced toxic effects. Variation in UGT1A1 has the potential to alter individual sensitivities to polycyclic aromatic hydrocarbons present in the diet and the environment.

Toxicological Relevance of UGTs (P.G.W., P.M.K.)

Toxicologic Implications of UGT Deficiencies. We have focused upon drugs and environmental chemicals for which toxicity depends upon the bioactivation of the xenobiotic or a stable metabolite by enzymes like the cytochromes P450 (P450s) or prostaglandin H synthases to highly toxic electrophilic and/or free radical reactive intermediates that damage cellular macromolecules (DNA, protein, lipid) and/or enhance oxidative stress (for reviews, see Wells and Winn, 1996; Wells et al., 1997). In such cases, glucuronidation and elimination of the xenobiotic and/or its stable metabolite serves as a toxicological gatekeeper, directing metabolism away from toxifying bioactivation. Furthermore, bioactivation often is a quantitatively minor pathway, often amounting to less than 10 to 15% of overall elimination, whereas glucuronidation usually is quantitatively major, constituting 40 to 75% or more of xenobiotic elimination. Hence, relatively minor deficiencies in glucuronidation theoretically can result in a substantial percentage increase in bioactivation, with toxicological consequences even at therapeutic drug concentrations or putatively safe concentrations of environmental chemicals.

Hepatic and Renal Toxicity. The widely used analgesic drug acetaminophen (paracetamol) at high doses is hepatotoxic and nephrotoxic, and eliminated primarily via UGT-catalyzed glucuronidation. Mutant Gunn and RHA rats, which lack all UGT1A enzymes (reviewed by Iyanagi et al., 1998), had reduced acetaminophen glucuronidation, enhanced P450-catalyzed bioactivation of acetaminophen and covalent binding to hepatocellular proteins, and enhanced hepatocellular centrilobular and renal cellular necrosis (de Morais and Wells, 1988, 1989; de Morais et al., 1992a). Interestingly, the ho-

mozygous UGT1A-deficient mutant rats exhibited measurable, albeit substantially reduced, acetaminophen glucuronidation, suggesting the likely contribution of a UGT2 isozyme. In people with a hereditary UGT1A1 deficiency (Gilbert's syndrome) (review: Tukey and Strassburg, 2000), similar studies using intravenous drug administration showed reduced glucuronidation of acetaminophen and, conversely, enhanced bioactivation determined by the formation of glutathione-derived acetaminophen metabolites, although measurable hepatotoxicity was not evident at the therapeutic dose employed (de Morais et al., 1992b). Among all the Gilbert's subjects and controls, a greater reduction in glucuronidation correlated highly with a greater enhancement in bioactivation. Conflicting data have been reported in other studies, with reduced acetaminophen glucuronidation apparent in some Gilbert's subjects and no effect in others (Ullrich et al., 1987; Esteban and Perez-Mateo, 1999). In our study, one subject with Gilbert's syndrome had normal acetaminophen glucuronidation, whereas one of the "normal" subjects without Gilbert's syndrome had deficient acetaminophen glucuronidation and enhanced bioactivation. These findings suggest that acetaminophen glucuronidation is more complex than previously thought, and may be explained by multiple UGT isozymes contributing to its glucuronidation, including UGT1A1 and 1A6. The extent to which each isozyme contributes may differ significantly among individuals, depending on exposure to inducing agents, and genetic and other factors. Interestingly, a recent study found that 8% of people with Gilbert's syndrome also possessed a homozygous deficiency in the UGT1A6 gene (Lampe et al., 1999).

Carcinogenesis. The environmental teratogens and/or carcinogens benzo[*a*]pyrene and 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) are representative of a multitude of polycyclic aromatic hydrocarbons and nitrosamines found widely in the environment, particularly in tobacco smoke. Most benzo[*a*]pyrene and NNK metabolites are eliminated primarily via glucuronidation, catalyzed by several UGT1A and UGT2 isozymes (Grove et al., 2000; Tukey and Strassburg, 2000), which avoids the alternative P450- and/or prostaglandin H synthase-catalyzed bioactivation of the metabolites to toxic reactive intermediates that irreversibly damage DNA (for reviews, see Wells and Winn, 1996; Wells et al., 1997). Using UGT1A-deficient Gunn and RHA rats, the glucuronidation of benzo[*a*]pyrene metabolites was found to be reduced in vitro and in vivo, resulting in their enhanced bioactivation and covalent binding to both protein and DNA (Hu and Wells, 1992). UGT deficiency was the critical toxicologic determinant, since these animals compared with controls had similar activities of various P450 isozymes and glutathione *S*-transferase isozymes and, in the case of the RHA strain, the controls were congenic (Hu and Wells, 1992). To estimate carcinogenic risk in UGT-deficient rats, a skin fibroblast model was developed, with genotoxicity and potential carcinogenicity assessed by the formation of micronuclei (Vienneau et al., 1995; Kim and Wells, 1996a). Cultured fibroblasts from UGT-deficient Gunn and RHA rats incubated with either benzo[*a*]pyrene or NNK had increased oxidative DNA damage and micronucleus formation compared with UGT-normal controls, whereas no increase in micronuclei occurred with benzo[*e*]pyrene, a noncarcinogenic isomer. Benzo[*a*]pyrene- and NNK-initiated micronucleus formation was dependent upon P450- and/or peroxidase-catalyzed bioactivation (Kim and Wells, 1996a; Kim et al., 1997a).

Cellular Models for Human Risk Assessment. To facilitate human studies, we have evaluated lymphocytes, which can be obtained in relatively large quantities sufficient for determination of collective UGT activity. To determine whether lymphocytes accurately reflect hepatic activities, lymphocytes and hepatic microsomes were taken/prepared from the same RHA UGT-normal (+/+) and UGT-deficient (j/j) rats (Hu and Wells, 1994). To produce benzo[*a*]pyrene metabolites as substrates for glucuronidation or ultimate bioactivation, benzo[*a*]pyrene was preincubated with rat liver microsomes and NADPH, and the supernatant was immediately added to the lymphocyte incubations. Lymphocytes from UGT-deficient rats accurately reflected the decreased glucuronidation of benzo[*a*]pyrene metabolites and enhanced bioactivation, covalent binding, and cytotoxicity that were observed with hepatic microsomes from the same rats (Hu and Wells, 1994), and with related in vivo studies of benzo[*a*]pyrene glucuronidation, bioactivation, and covalent binding (Hu and Wells, 1992), and embryotoxicity (Wells et al., 1989). Lymphocytes may constitute a useful model for risk assessment.

In preliminary human studies, lymphocytes from 12 normal volunteers were tested as described in the rat lymphocyte studies above (Hu and Wells, 1993). All subjects had normal UGT activity for bilirubin (i.e., none had Gilbert's syndrome), but there was a 200-fold variability in UGT activities for benzo[*a*]pyrene metabolites, including two with no measurable activity. Decreasing UGT activity correlated with decreased UGT-dependent protection against benzo[*a*]pyrene covalent binding and, conversely, with increased cytotoxicity for benzo[*a*]pyrene quinones and diols, but not monophenols. A similar protection against benzo[*a*]pyrene quinones, but not 3-OH-benzo[*a*]pyrene or benzo[*a*]pyrene-7,8-dihydrodiol, was shown in human lymphoblastoid cells transfected with rat UGT1A7 (Grove et al., 2000). Our lymphocyte studies suggest that substantial UGT deficiencies for potentially toxic benzo[*a*]pyrene metabolites are common in the normal population and indicate that these UGT deficiencies may constitute important determinants of toxicologic predisposition, particularly with respect to chemical carcinogenesis and teratogenesis.

Developmental Toxicity. Whereas there is little UGT activity in the rodent embryo, maternal UGT activity theoretically may play an important role in determining how much teratogen reaches the embryo. In preliminary studies, pregnant UGT-deficient Gunn rats were substantially more susceptible to benzo[*a*]pyrene-initiated embryonic death at a subcarcinogenic dose (25 mg/kg i.p.) that had no effect on UGT-normal Wistar controls (Wells et al., 1989). Similarly, the anticonvulsant drug phenytoin, a human teratogen, and its major metabolite, 5-(*p*-hydroxyphenyl)-5-phenylhydantoin, caused increased DNA oxidation and micronucleus formation in UGT-deficient cultured rat skin fibroblasts, and in vivo studies indicated that this enhanced genotoxicity was due to decreased *N*-glucuronidation of phenytoin and *O*-glucuronidation of 5-(*p*-hydroxyphenyl)-5-phenylhydantoin in the UGT-deficient rats (Kim et al., 1997b), resulting in increased hydroxyl radical formation (Kim and Wells, 1996b). Preliminary in vivo evidence suggests a similar enhancement in embryopathies in phenytoin-treated pregnant Gunn and RHA UGT-deficient rats treated with phenytoin (Kim and Wells, 1998). In the last trimester of pregnancy, increasing activities of some fetal UGTs may provide the offspring with a second layer of biochemical protection, particularly against functional (as distinct from structural) anomalies, although this has yet to be studied.

General Toxicologic Observations. Overall, several consistent observations emerged from the above studies. First, a relatively small percentage decrease in a quantitatively major pathway of elimination, such as glucuronidation, can produce a disproportionately large percentage increase in bioactivation. This effect is particularly evident if there are no alternative eliminating pathways, or if the alternative pathways are readily saturable, as is the case with sulfation. For example, even in the rat, which has at least twice the sulfating capacity of mice and humans for acetaminophen, there was no compensatory enhancement of acetaminophen sulfation in UGT-deficient

rats (de Morais et al., 1992a). The toxicological consequences in humans may be more pronounced.

Second, consistent, progressively greater deficiencies in UGT activity in +/j and j/j RHA rats resulted in corresponding *UGT* gene dose-dependent decreases in the glucuronidation of both acetaminophen and benzo[*a*]pyrene in vitro and/or in vivo and, conversely, increasing xenobiotic bioactivation, covalent binding, and toxicity, all of which were reflected in the lymphocyte model. A consistent finding of particular clinical interest is that a heterozygous deficiency in UGTs was toxicologically relevant (de Morais et al., 1992a; Hu and Wells, 1992, 1994), with a risk equivalent to homozygotes in some systems (Kim and Wells, 1996b; Kim et al., 1997b).

Finally, despite our improving understanding of the pharmacogenomic basis for UGT deficiencies and the specific roles of UGT isozymes in xenobiotic glucuronidation (reviews: Tukey and Strassburg, 2000; Guillemette, 2003), the complexity of toxicological risk remains challenging. At a given xenobiotic concentration, individual toxicological susceptibility will depend upon both the levels of the relevant UGT proteins, which vary according to genetic and environmental determinants, and the overall balance among numerous associated biochemical pathways, including elimination via other conjugating pathways, membrane transport, bioactivation, reactive intermediate detoxification, and macromolecular repair, among others. Environmental modulation of the in vivo outcome can be particularly unpredictable, as exemplified by acetaminophen toxicity, which in rodents is reduced by the UGT inducer oltipraz due to enhanced glucuronidation (Davies and Schnell, 1991; Kessler et al., 2002), but conversely is enhanced by pretreatment with the UGT inducers phenobarbital and 3-methylcholanthrene, presumably due to their relatively greater induction of P450-catalyzed bioactivation.

Pharmacogenetics of UGT Enzymes: Implications for Cancer Susceptibility (C.G.)

The molecular genetics of the UGT superfamily are well understood (Mackenzie et al., 1997; Gong et al., 2001), but the molecular mechanisms of large interindividual phenotypic variations remain to be elucidated. Breakthroughs in the identification of functional common genetic variations in *UGT* genes that may impact drug response and/or disease susceptibility are starting to emerge. Polymorphic variations that affect functional activity of the enzyme have been characterized in human *UGT1A* and *UGT2B* genes. Similar to other drug-metabolizing enzymes, evidence of differential prevalence among ethnic and racial groups have been observed in human populations (Fig. 2) (reviewed in Guillemette, 2003).

The first evidence suggesting alterations in *UGT* genes as a genetic risk factor of cancer was recently obtained. In a first study, we hypothesized that constitutive alteration in UGTs involved in the inactivation of estradiol and its catechol-reactive metabolites may modify estrogen exposure and, consequently, estrogen-related cancer risk. We investigated UGT1A1 as a first candidate gene. UGT1A1 is a major UGT expressed in mammary gland and involved in the formation of estradiol-glucuronide (Senafi et al., 1994; Guillemette et al., 2000a). The most common genetic variant in the *UGT1A1* gene is a dinucleotide repeat polymorphism in the atypical TATA-box region of the UGT1A1 promoter (Bosma et al., 1995). Using genetic epidemiological studies designed as population-based case-control studies, we first observed that the UGT1A1*28 (A(TA)7TAA) and the UGT1A1*34 (A(TA)8TAA) promoter alleles were associated with an increased risk of developing invasive breast cancer in premenopausal African-American women (OR = 2.1; 95% confidence interval, 1.0–4.2; $p = 0.04$) (Guillemette et al., 2000a). This finding is consistent with a role for estrogen-UGT in modulating the action of endogenous hormones in breast cancer risk in the African-American population. However, in a nested case-control study of white women within the Nurses' Health Study cohort, we were unable to detect a significant risk associated with the low transcriptional allele UGT1A1*28 (Guillemette et al., 2001). In the same population, an elevation of estradiol among women who are carriers of at least one UGT1A1*28 allele was observed and suggests a possible contribution of the glucuronidation pathway, and especially UGT1A1, in the maintenance of hormone homeostasis, although not sufficient to alter breast cancer risk in white women. We further studied the relationship between UGT1A1 polymorphisms and variation in breast density, a predictor of breast cancer. Premenopausal women homozygous for the UGT1A1*28 allele were found to have significantly lower breast density compared with those with the *1/*1 genotype (−43.1% difference; $p = 0.04$). In contrast, postmenopausal women with the UGT1A1*28/*28 genotype had greater breast density compared with those with the *1/*1 genotype (+32.0% difference; $p = 0.05$), which was even greater among current postmenopausal hormone users (+56.8% difference; $p = 0.03$). These results suggest that UGT1A1 genotype is a predictor of breast density within groups of different menopausal status and support that interindividual differences in estrogen glucuronidation influence local estrogen concentration and breast density (Haiman et al., 2003).

Recent studies support the idea that UGTs play a significant role in the detoxification of environmental carcinogens and therefore represent good candidates for low-penetrance susceptibility genes that possibly contribute to cancer risk by amplifying the effects of carcinogen exposure. Among these, UGT1A7 is an important extrahepatic UGT and represents one of several UGTs that were shown to be active on chemical carcinogens. In a recent study, we identified three common UGT1A7 allelic variant isozymes that encode novel UGT1A7 proteins differing in primary structure at three amino acid positions (Fig. 2) (Guillemette et al., 2000b; Woolley et al., 2000). Prevalence studies confirmed the presence of all four UGT1A7 alleles in the African-American and Asian populations, although at various frequencies. Functional studies revealed significant differences in catalytic activity of the UGT1A7 alleles toward benzo(*a*)pyrene metabolites, known tobacco-carcinogens, as well as for a number of additional substrates (Guillemette et al., 2000b). UGT1A7 was shown to be expressed in normal orolaryngeal tissue specimens including tongue, tonsil, floor of mouth, and larynx, which are target tissues of carcinogenic environmental molecules (Zheng et al., 2001). The UGT1A7 allele, associated with low activity, was subsequently associated with an increased risk of orolaryngeal cancer (Zheng et al., 2001). Genotypes containing different combinations of the two lowest activity alleles, UGT1A7*3 ($129^{Lys}131^{Lys}208^{Arg}$) and UGT1A7*4 ($129^{Arg}131^{Asn}208^{Arg}$), were strongly linked to an increased risk for orolaryngeal cancer in white (OR = 2.8; 95% IC 1.1–7.6) and in African-American people (OR = 6.2; 95% IC 1.2–31) compared with the wild-type genotype (homozygous for the UG1A7*1 allele [$129^{Arg}131^{Asn}208^{Trp}$]). Upon stratification by cancer site, predicted low-activity UGT1A7 genotypes were strongly linked to increased risk for both oral cavity (OR = 4.2; 95% IC 1.7–0) and laryngeal cancers (OR = 3.7; 95% IC 0.99–14). In addition, subjects with low-activity genotypes who were light or heavy smokers had a significantly increased risk compared with the wild-type genotype (OR = 3.7; 95% IC 1.1–12 and OR = 6.1; 95% IC 1.5–25, respectively) (Zheng et al., 2001). These results revealed for the first time that genetic variations in the *UGT1A7* gene that reduce the carcinogen-detoxifying activity increase the risk of developing a smoking-related orolaryngeal cancer. The association of UGT1A7 alleles with risk of hepatic and colorectal cancers was further demonstrated (Vo-

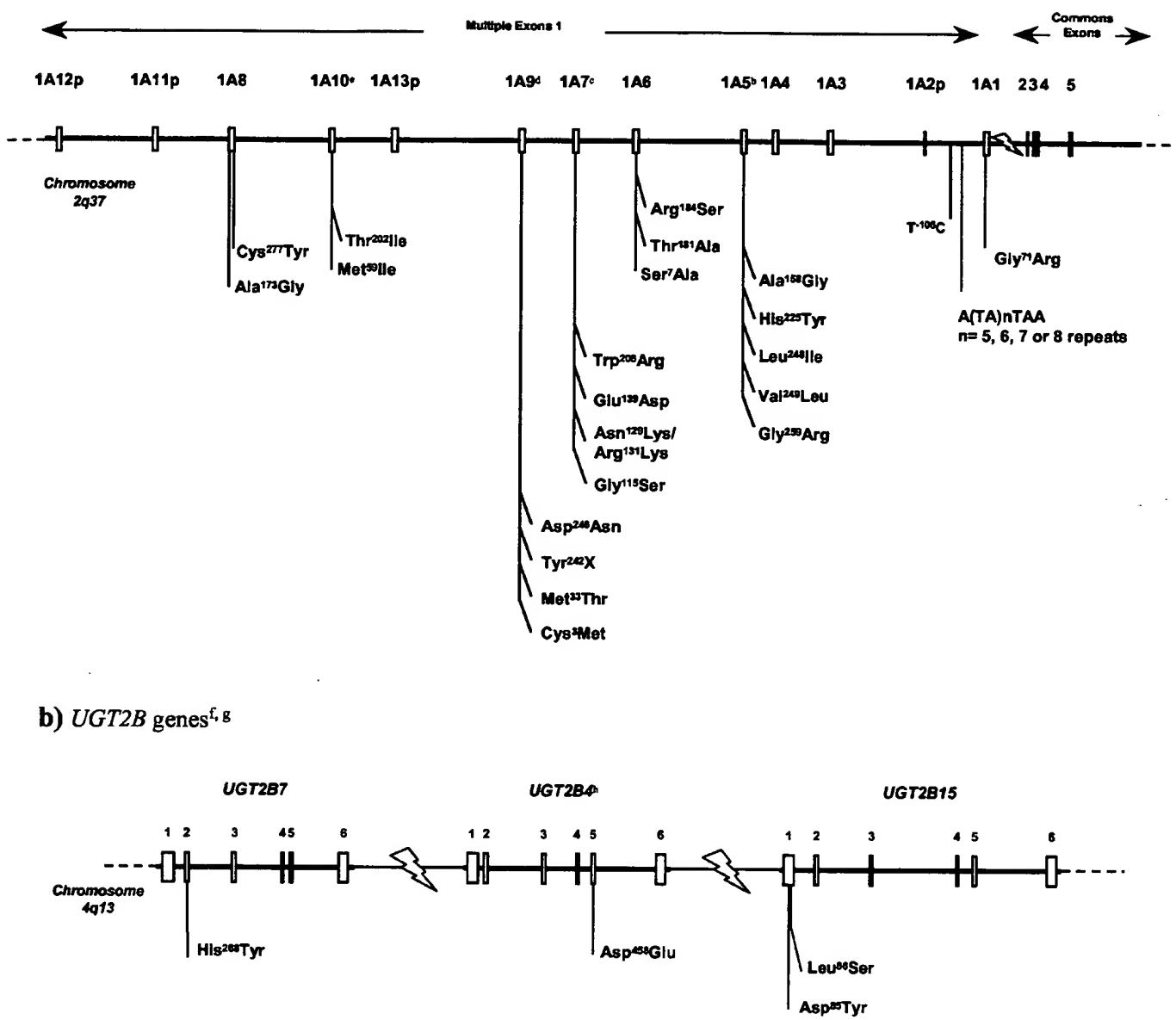


FIG. 2. *Common genetic polymorphisms in UGT1A and UGT2B genes.*

A, common genetic polymorphisms at the *UGT1A* gene locus. The entire UGT1 family is derived from a single gene locus (*UGT1*), located on chromosome 2q37, which encodes nine functional proteins (UGT1A1, and UGT1A3 to UGT1A10) and four pseudogenes (UGT1A2p, and UGT1A11p to UGT1A13p) (Ritter et al., 1992; Gong et al., 2001). B, common genetic polymorphisms in functional *UGT2B* genes. The UGT2B family comprises several distinct genes and pseudogenes, which are not included here. The genomic organization of functional *UGT2B* genes is not entirely elucidated; therefore, the relative chromosomal localization used here was simplified for schematic purposes. Sequences that differ by less that 3% are considered alleles of the same gene (Mackenzie et al., 1997). To date, allelic variants have been reported for UGT1A1, UGT1A6, UGT1A7, UGT1A8, and UGT1A9, in addition to UGT2B4, UGT2B7, UGT2B15, and UGT2B28. The function and prevalence of some of these variants have also been described (Guillemette et al., 2000, 2001; Woolley et al., 2000; Girard et al., 2003; Beutler et al., 1998; Bosma et al., 1994; Ciotti et al., 1997; Coffman et al., 1998; Lampe et al., 1999, 2000; Guillemette et al., 2000; Hall et al., 1999; Riedy et al., 2000; Lévesque et al., 1997, 1999; Turgeon et al., 2000; Akaba et al., 1998; Maruo et al., 1999; Huang et al., 2002; Villeneuve et al., 2003).[A] Recent studies support a possible role of UGT in cancer risk (Guillemette et al., 2000, 2001; Zheng et al., 2001; Vogel et al., 2001; Strassburg et al., 2002; Gsur et al., 2002; MacLeod et al., 2000; Gestl et al., 2002), although additional studies are needed to confirm these findings.[B]. [a] The structure of the *UGT1* gene presented here is based on the GenBank accession number AF297093 (Gong et al., 2001). [b] UGT1A5 alleles correspond to: UGT1A5*1 $Ala^{158}His^{225}Leu^{248}Val^{249}Gly^{259}$; UGT1A5*2 $Gly^{158}His^{225}Leu^{248}Val^{249}Gly^{259}$; UGT1A5*3 $Ala^{158}Tyr^{225}Leu^{248}Val^{249}Gly^{259}$; UGT1A5*4 $Ala^{158}His^{225}Ile^{248}Leu^{249}Arg^{259}$; UGT1A5*5 $Gly^{158}His^{225}Ile^{248}Leu^{249}Arg^{259}$; UGT1A5*6 $Ala^{158}Tyr^{225}Ile^{248}Leu^{249}Arg^{259}$; UGT1A5*7 $Gly^{158}Tyr^{225}Ile^{248}Leu^{249}Arg^{259}$. [c] UGT1A7 alleles correspond to: UGT1A7*1 $Gly^{115}Asn^{129}Arg^{131}Glu^{139}Trp^{208}$; UGT1A7*2 $Gly^{115}Lys^{129}Lys^{131}Glu^{139}Trp^{208}$; UGT1A7*3 $Gly^{115}Lys^{129}Lys^{131}Glu^{139}Arg^{208}$; UGT1A7*4 $Gly^{115}Asn^{129}Arg^{131}Glu^{139}Arg^{208}$; UGT1A7*5 $Ser^{115}Asn^{129}Arg^{131}Glu^{139}Arg^{208}$; UGT1A7*6 $Gly^{115}Asn^{129}Arg^{131}Glu^{139}Trp^{208}$; UGT1A7*7 $Gly^{115}Lys^{129}Lys^{131}Asp^{139}Trp^{208}$ UGT1A7*8 $Gly^{115}Lys^{129}Lys^{131}Asp^{139}Arg^{208}$ and UGT1A7*9 $Ser^{115}Lys^{129}Lys^{131}Glu^{139}Tr^{208}$. [d] UGT1A9 alleles correspond to: UGT1A9*1 $Cys^{3}Met^{33}$; UGT1A9*2 $Tyr^{3}Met^{33}$; UGT1A9*3 $Cys^{3}Thr^{33}$ (Villeneuve et al., 2003); UGT1A9*4 $Tyr^{242}X$ (Y. Saito, unpublished data), and UGT1A9*5 $Asp^{246}Asn$ (Jinno et al., 2003a). [e] UGT1A10 alleles correspond to: UGT1A10*1 $Met^{59}Thr^{202}$; UGT1A10*2 $Ile^{59}Thr^{202}$; and UGT1A10*3 $Met^{59}Ile^{202}$ (Jinno et al., 2003b). [f] The relative positions of the *UGT2B4*, *UGT2B7*, and *UGT2B15* genes on chromosome 4q13 are based on the data reported by Riedy et al. (2000). [g] Polymorphic expression of two truncated UGT2B28 variants (type II and type III) has been reported (Levesque et al., 2001). UGT2B28 type II differs from type I by a deletion of 308 bp in the cofactor binding domain, whereas UGT2B28 type III lacks 351 bp in the putative substrate binding domain. [h] An additional cDNA clone isolated from human liver corresponds to the UGT2B4 $Phe^{109}Leu$, $Phe^{396}Leu$ allele but appears to be very rare since it was not found in two independent population studies.

gel et al., 2001; Zheng et al., 2001; Strassburg et al., 2002). Recently, we conducted a case-control study, 400 cases and 400 controls matched for age, sex, and race, to assess the relation between characteristics of meat consumption, heterocyclic amine (HCA) exposure, the UGT1A7 genotype, and colon cancer. No main effect of the UGT1A7 genotype was observed on colon cancer risk. On the other hand, the association between dietary HCA exposure and colon cancer was modified in individuals with the low-activity UGT1A7 genotypes (C. Guillemette, unpublished observations). These data suggest that the relation between dietary sources of HCA and colon cancer may be modulated by the UGT1A7 detoxification pathway. These results also point to HCA exposure as an important etiologic factor in colon cancer. Altogether, these findings warrant additional epidemiological studies to confirm the role of low UGT1A7 conjugator genotypes in risk for cancers. Besides, our group recently discovered two additional *UGT1A7* single nucleotide polymorphisms, found exclusively in African-American subjects, which generate five additional alleles (UGT1A7*5 to *9) when combined with the four known single nucleotide polymorphisms present in UGT1A7*2, *3, and *4. Upon functional analysis, several of these UGT1A7 variant isozymes exhibited much lower glucuronidation activities compared with UGT1A7*1; their role in cancer remains unverified at the present time (Girard et al., 2003).

In conclusion, much research in this area is needed, although promising leads have emerged regarding the role of UGT genetic polymorphisms in cancer etiology and on their possible implication in modulating the degree of exposure to several carcinogenic compounds.

Gene Therapy for UGT Deficiencies (J.R.C., N.R.C.)

Genetic lesions of *UGT1A1* can result in three grades of hyperbilirubinemia in humans. Insertions, deletions, or mutations of any of the five exons encoding UGT1A1 that cause near-complete loss of the enzyme activity result in the potentially lethal disorder, Crigler-Najjar syndrome type 1 (CN-1). Mutations causing lesser degrees of reduction of UGT1A1 activity cause a milder form of the disease, termed Crigler-Najjar syndrome type 2. The third grade of hyperbilirubinemia is an even milder form, termed Gilbert's syndrome, which results from the insertion of TA dinucleotides within the TATAA element of the UGT1A1 promoter or base substitutions in the UGT1A1 coding region (Roy Chowdhury et al., 2001). Of these three disorders, CN-1 is associated with levels of unconjugated hyperbilirubinemia that are severe enough to cause bilirubin encephalopathy. Before the routine use of phototherapy, CN-1 was generally lethal during infancy. Although phototherapy permits survival beyond adolescence, it becomes progressively less effective around puberty, so that the risk of kernicterus persists life-long. Currently, liver transplantation is the only definitive therapy for CN-1. However, discovery of the molecular bases of inherited jaundice, and advances in the techniques of hepatocyte transplantation and nucleic acid transfer to the liver have brought gene therapy for Crigler-Najjar syndrome close to reality. Gene therapy methods can be classified into approaches based on isolated hepatocyte, and methods using gene delivery in vivo.

Hepatocyte-Based Gene Therapy. In the simplest form, normal genes may be introduced into the liver of patients by transplantation of allogeneic normal human hepatocytes. Hepatocytes introduced into the portal circulation by infusion into the portal vein or injection into the splenic pulp integrate into normal liver chords of structurally normal liver with remarkable rapidity and function on a long-term basis (Gupta et al., 1991). Immunosuppression is required for prevention of allograft rejection. This method should be particularly useful for diseases, such as CN-1, in which the liver architecture is normal. Because the host liver remains intact, the metabolic cost of graft rejection is limited. A 10-year-old girl with CN-1 was the first to receive liver transplantation (Fox et al., 1998). Introduction of 7.5×10^9 normal isolated hepatocytes into the liver through a percutaneously placed portal vein catheter reduced serum bilirubin to approximately half the pretransplant level for over 2 years. This study demonstrated the safety and efficacy of hepatocyte transplantation, but despite the replacement of approximately 5% of the hepatic UGT1A1 activity, the need for phototherapy was not obviated (Roy Chowdhury et al., 1998). It appears that complete cure of metabolic liver diseases will require repeated hepatocyte transplantation or preferential proliferation of the transplanted hepatocytes over the host cells.

To solve many of the lingering problems associated with hepatocyte transplantation, extensive studies are ongoing in Gunn rats that are both a molecular and pathophysiological model of CN-1. Such preferential proliferation requires a strong mitotic stimulus to the liver, to which only the engrafted cells, but not the host hepatocytes, can respond. This was achieved by the combination of preparative irradiation of the liver and partial hepatectomy, prior to hepatocyte transplantation (Guha et al., 1999, 2002). Controlled irradiation of the liver prevents proliferation of the host Gunn rat hepatocytes. Consequently, transplanted congeneic normal hepatocytes progressively repopulate the liver almost completely by 12 weeks, fully normalizing serum bilirubin levels.

Another approach involves isolating hepatocytes from a resected liver segment, transducing the primary hepatocytes with a therapeutic gene in culture and subsequently transplanting the cells back into the donor (Roy Chowdhury et al., 1991). Because the cells are autologous, immune suppression is not needed. However, the number of cells that can be harvested, transduced, and engrafted after transplantation is limited, which severely restricts the efficiency of ex vivo gene therapy. Some of these limitations may be overcome by conditionally immortalizing the hepatocytes, so that the transduced cells can be expanded in culture before transplantation (Tada et al., 1998).

Gene Transfer in Vivo. The conventional in vivo gene therapy methods consist of replacing a missing functional gene by transferring nucleic acids to the target cells using viral or nonviral vectors. A radically new approach involves repairing genetic mutations in situ. These approaches, as related to the treatment of UGT1A1 deficiency, are briefly described below.

Nonviral vectors. Physical methods, such as ballistics or direct injection into the liver parenchyma, have resulted in a limited extent of gene transfer. Naked DNA can be transfected into the liver by rapid intravenous administration of plasmids at high volumes, causing a volume overload and hepatic congestion. Obviously this approach is not translatable to clinical application. DNA can be complexed electrostatically with polycations, which can be lactosylated or complexed with galactose-terminated peptides for hepatocyte-targeted delivery in vivo. The ligand-DNA complex is internalized by receptor-mediated endocytosis via hepatocyte-specific asialoglycoprotein receptors. Although the majority of the ligand is translocated to the lysosome, a small fraction reaches the nucleus, where it is expressed transiently. The transgene expression can be prolonged to several weeks by performing partial hepatectomy or transient pharmacological disruption of the microtubules (Roy Chowdhury et al., 1996) or by using polycations, such as polyethyleneimine, that destabilize the endosomal vesicles. An important new advance in plasmid-directed gene transfer has been the use of a Tc1 Mariner-type transposon system, termed the *Sleeping Beauty*. Expression of the *Sleeping Beauty* transposase results in the transposition of stretches of DNA, flanked by inverted/direct repeats of the transposon sequences, into host cellular

chromosomes, leading to long-term transgene expression (Izsvák et al., 2000).

Viral vectors. Recombinant viruses capable of transferring genes into cells in vitro or in vivo are commonly used for gene therapy. These vectors can be classified into those that remain episomally and those that integrate into the host genome. Recombinant adenoviruses are prototypes of episomal vectors. Adenoviral vectors transfer genes into nondividing hepatocytes in vivo with great efficiency, although the longevity of the episomal vector is limited. Moreover, adenoviral proteins are strongly immunogenic, and host humoral and cellular immune response precludes repeated gene transfer. Specific tolerization of the host by administration in neonatal rats (Takahashi et al., 1996), intrathymic injection of adenoviral proteins (Ilan et al., 1996), or oral administration of small doses of adenoviral proteins (Ilan et al., 1997) have resulted in specific host tolerance to the viral proteins and have permitted repeated administration of the recombinant virus and long-term amelioration of jaundice in Gunn rats. Recently, immunomodulatory genes, such as CTLA4Ig, have been incorporated into adenoviral vectors (Thummala et al., 2002). Coexpression of such genes, along with the transgene, makes the virus nonimmunogenic, prolongs the transgene expression, and permits repeated administration.

Retroviral vectors integrate into the host genome. Although recombinant oncoretroviruses have been used extensively for gene therapy (Roy Chowdhury et al., 1991; Tada et al., 1998), these vectors require host cell mitosis for integration, which is infrequent in the quiescent liver. Vectors based on immunodeficiency-type retroviruses (lentiviruses) that can integrate into nondividing cells are being developed for the treatment of UGT1A1 deficiency. Recombinant adenoassociated viruses (rAAVs) are also being tested in Gunn rats, although, so far, it has been possible to transduce only up to 5% of rat hepatocytes by intraportal infusion. Recent studies indicate that rAAVs do not integrate to a significant extent in immunocompetent animals, suggesting that, in contrast to previous expectations, the transgene expression may not be permanent (Nakai et al., 2001). Moreover, because rAAV vectors evoke a host antibody response, repeated administration may be problematic. Therefore, the search for newer vectors continues. Recombinant simian virus 40 is promising as an integrating vector because this T-antigen-deleted virus does not evoke host immune response. Recombinant simian virus 40 integrates into the host genome progressively over several days. After infusion of a recombinant SV40 expressing human UGT1A1 into the portal vein of the Gunn rat, serum bilirubin was reduced by up to 60% and remained at that level throughout an 18-month period of observation, suggesting a permanent therapeutic effect. There was no detectable antibody response to the recombinant virus, and gene transfer was repeatable upon injection of a recombinant SV40 vector expressing a different transgene (Sauter et al., 2000). These viruses can be grown and concentrated to infectious titers of 10^{11} to 10^{12}. Thus, the recombinant SV40 appears to have a great potential in liver-directed gene therapy, its major limitation being a relatively small DNA packaging space (4–4.5 kilobases) (Strayer et al., 2002).

Gene repair therapy. Site-directed gene repair in vivo is a novel form of gene therapy, which relies on the cellular DNA repair mechanisms to correct point mutations or short deletions (Kren et al., 1999). The method utilizes synthetic RNA-DNA chimera, which align to a target sequence in the genome with high specificity and efficiency. The nucleotide sequence in the chimera is complementary to the target genomic sequence, except for a single mismatch. After alignment, the mismatch between the DNA limb of the chimeric molecule and the complementary strand of the genomic DNA triggers the cell's mismatch repair enzymes, resulting in correction of the mutation. For correction of the single guanosine base deletion in the *UGT1A1* exon 4 of Gunn rats, the RNA-DNA chimera was constructed to contain the wild-type sequence (Kren et al., 1999). The chimeric molecules were complexed with lactosylated polyethyleneimine or with galactosylated liposomes. After intravenous infusion of the carrier-chimera complex, the genetic lesions were corrected in 1 to 10% of the alleles, resulting in the appearance of the enzymatically active full-length UGT1A1 in the liver of the recipient Gunn rats. Function of the enzyme was demonstrated by the appearance of bilirubin glucuronides in bile and reduction of plasma bilirubin levels to 35 to 40% of pretreatment levels. Long-term follow-up indicates permanent correction of the genetic lesion. Figure 3 shows immunohistochemical staining of liver sections from Gunn rats treated with recombinant adenoviral or SV40 vectors, or with synthetic DNA-RNA chimera for gene repair. In summary, both nonviral and viral vectors are being rapidly improved for overcoming the challenges of liver-directed gene therapy for CN-1. Hepatocyte transplantation and gene therapy are being developed in parallel to complement each other.

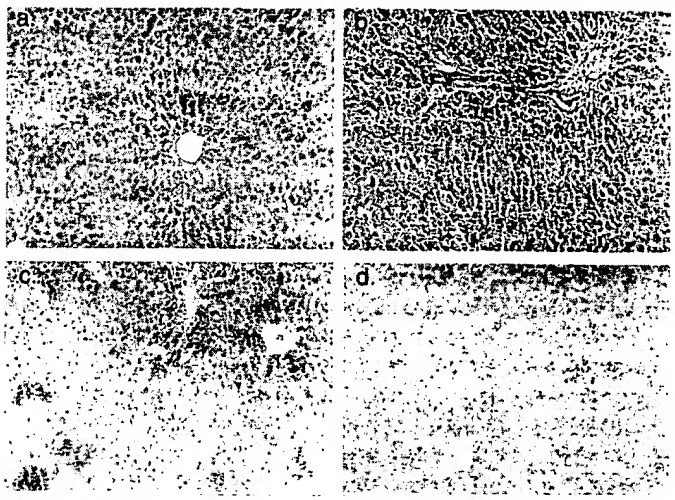


FIG. 3. *Immunohistochemical staining of sections of Gunn rat liver with anti-UGT1A1 antibodies.*

For panels a–c, immunohistochemical staining was performed using a human UGT1A-specific monoclonal antibody. Panel a, control (untreated) Gunn rat; panel b, liver from a Gunn rat 1 week after the injection of a first-generation adenovirus expressing human UGT1A1 (10^9 pfu). Panel c, liver from a Gunn rat 1 year after the injection of 10^8 pfu of a recombinant SV40 expressing human UGT1A1. Panel d, gene repair therapy in Gunn rat. Liver biopsy was performed 1 year after five injections of an RNA-DNA chimera designed to insert the missing guanosine residue in Gunn rat *UGT1A1* (exon 4). Immunohistochemical staining was performed using a rabbit antiserum specific for rat UGT1A1. Please note that the staining of the positive cells is light, probably indicating that only one allele per cell is corrected.

It is apparent that significant progress continues to be made in our understanding of the UGT gene family and their roles in physiology and pathophysiology. Molecular factors involved in the control of expression of the UGTs are being identified, providing clues to how these genes are regulated from tissue to tissue, during development, and in response to certain chemical exposures. The substrate specificities and activities of the different UGT family members in humans are beginning to be understood, and information on their rodent counterparts is now coming to light which will be valuable for interpretation of toxicity studies in rodents. Identification of genetic differences among individuals in the sequences of UGT genes and their impact on risks of toxicity from exposure to drugs and environmental chemicals is a particularly active area of investigation and will be useful for predictive risk assessment. The research comes full circle

with current efforts toward development of an effective gene therapy for treatment of Crigler-Najjar syndrome, providing a greatly needed alternative to liver transplantation. It should be clear that advances from both basic and clinical research involving animals and humans have provided critical insights into the role of glucuronidation and UGTs in health, drug therapy, and disease.

References

Akaba K, Kimura T, Sasaki A, Tanabe S, Ikegami T, Hashimoto M, Umeda H, Yoshida H, Umetsu K, Chiba H, et al. (1998) Neonatal hyperbilirubinemia and mutation of the bilirubin uridine diphosphate-glucuronosyltransferase gene: a common missense mutation among Japanese, Koreans and Chinese. *Biochem Mol Biol Int* **46**:21–26.

Beaulieu M, Levesque E, Tchernof A, Beatty BG, Belanger A, and Hum DW (1997) Chromosomal localization, structure and regulation of the UGT2B17 gene, encoding a C19 steroid metabolizing enzyme. *DNA Cell Biol* **16**:1143–1154.

Bernard P, Goudonnet H, Artur Y, Desvergne B, and Wahli W (1999) Activation of the mouse TATA-less and human TATA-containing UDP-glucuronosyltransferase 1A1 promoters by hepatocyte nuclear factor 1. *Mol Pharmacol* **56**:526–536.

Beutler E, Gelbart T, and Demina A (1998) Racial variability in the UDP-glucuronosyltransferase 1 (UGT1A1) promoter: a balanced polymorphism for regulation of bilirubin metabolism? *Proc Natl Acad Sci USA* **95**:8170–8174.

Bock KW, Forster A, Gschaidmeier H, Bruck M, Munzel P, Schareck W, Fournel-Gigleux S, and Burchell B (1993) Paracetamol glucuronidation by recombinant rat and human phenol UDP-glucuronosyltransferases. *Biochem Pharmacol* **45**:1809–1814.

Bosma PJ, Roy Chowdhury J, and Bakker C (1995) The genetic basis of the reduced expression of bilirubin UDP-glucuronosyltransferase 1 in Gilbert's syndrome. *New Engl J Med* **333**: 1171–1175.

Bosma PJ, Seppen J, Goldhoorn B, Bakker C, Oude Elferink RP, Roy Chowdhury J, Roy Chowdhury N, and Jansen PL (1994) Bilirubin UDP-glucuronosyltransferase 1 is the only relevant bilirubin glucuronidating isoform in man. *J Biol Chem* **269**:17960–17964.

Ciotti M, Marrone A, Potter C, and Owens IS (1997) Genetic polymorphism in the human UGT1A6 (planar phenol) UDP-glucuronosyltransferase: pharmacological implications. *Pharmacogenetics* **7**:485–495.

Coffman BL, King CD, Rios GR, and Tephly TR (1998) The glucuronidation of opioids, other xenobiotics, and androgens by human UGT2B7Y(268) and UGT2B7H(268). *Drug Metab Dispos* **26**:73–77.

Court MH, Duan SX, von Moltke LL, Greenblatt DJ, Patten CJ, Miners JO, and Mackenzie PI (2001) Interindividual variability in acetaminophen glucuronidation by human liver microsomes: identification of relevant acetaminophen UDP-glucuronosyltransferase isoforms. *J Pharmacol Exp Ther* **299**:998–1006.

Davies MH and Schnell RC (1991) Oltipraz-induced amelioration of acetaminophen hepatotoxicity in hamsters. II. Competitive shunt in metabolism via glucuronidation. *Toxicol Appl Pharmacol* **109**:29–40.

de Morais SMF, Chow SYM, and Wells PG (1992a) Biotransformation and toxicity of acetaminophen in congenic RHA rats with or without a hereditary deficiency in UDP-glucuronosyltransferase. *Toxicol Appl Pharmacol* **117**:81–87.

de Morais SMF, Uetrecht JP, and Wells PG (1992b) Decreased glucuronidation and increased bioactivation of acetaminophen in Gilbert's syndrome. *Gastroenterology* **102**:577–586.

de Morais SMF and Wells PG (1988) Deficiency in bilirubin UDP-glucuronyl transferase as a genetic determinant of acetaminophen toxicity. *J Pharmacol Exp Ther* **247**:323–331.

de Morais SMF and Wells PG (1989) Enhanced acetaminophen toxicity in rats with bilirubin UDP-glucuronyl transferase deficiency. *Hepatology* **10**:163–167.

Ebner T, Remmel RP, and Burchell B (1993) Human bilirubin UDP-glucuronosyltransferase catalyzes the glucuronidation of ethinylestradiol. *Mol Pharmacol* **43**:649–654.

Esteban A and Perez-Mateo M (1999) Heterogeneity of paracetamol metabolism in Gilbert's syndrome. *Eur J Drug Metab Pharmacokinet* **24**:9–13.

Fang JL, Beland FA, Doerge DR, Wiener D, Guillemette C, Marques MM, and Lazarus P (2002) Characterization of benzo(a)pyrene-trans-7,8-dihydrodiol glucuronidation by human tissue microsomes and overexpressed UDP-glucuronosyltransferase enzymes. *Cancer Res* **62**:1978–1986.

Fox IJ, Roy Chowdhury J, Kaufman SS, Goertzen TC, Chowdhury NR, Warkentin PI, Dorko K, Sauter BV, and Strom SC (1998) Treatment of the Crigler-Najjar syndrome type 1 with hepatocyte transplantation. *N Engl J Med* **338**:1422–1426.

Gestl SA, Green MD, Shearer DA, Frauenhoffer E, Tephly TR, and J Weisz (2002) Expression of UGT2B7, a UDP-glucuronosyltransferase implicated in the metabolism of 4-hydroxyestrone and all-trans retinoic acid, in normal human breast parenchyma and in invasive and in situ breast cancers. *Am J Pathol* **160**:1467–1479.

Girard H, Journault K, and Guillemette C (2003) Haplotypic structure of the carcinogen-metabolizing enzyme UGT1A7: nine polymorphic alleles (*1 through *9). Abstract 7485, in *ASPET, Experimental Biology 2003*, April 11–15, 2003. San Diego, CA.

Gong QH, Cho JW, Huang T, Potter C, Gholami N, Basu NK, Kubota S, Carvalho S, Pennington MW, Owens IS, and Popescu NC (2001) Thirteen UDPglucuronosyltransferase genes are encoded at the human UGT1 gene complex locus. *Pharmacogenetics* **11**:357–368.

Gregory PA, Hansen AJ, and Mackenzie PI (2000) Tissue specific differences in the regulation of the UDP glucuronosyltransferase 2B17 gene promoter. *Pharmacogenetics* **10**:809–820.

Grove AD, Kessler FK, Metz RP, and Ritter JK (1997) Identification of a rat oltipraz-inducible UDP-glucuronosyltransferase (UGT1A7) with activity towards benzo(a)pyrene-7,8-dihydrodiol. *J Biol Chem* **272**:1621–1627.

Grove AD, Llewellyn GC, Kessler FK, White KL, Crespi CL, and Ritter JK (2000) Differential protection by rat UDP-glucuronosyltransferase 1A7 against benzo[a]pyrene-3,6-quinone- versus benzo[a]pyrene-induced cytotoxic effects in human lymphoblastoid cells. *Toxicol Appl Pharmacol* **162**:34–43.

Gsur A, Preyer M, Haidinger G, Schatzl G, Madersbacher S, Marberger M, Vutuc C, and Micksche M (2002) A polymorphism in the UDP-glucuronosyltransferase 2B15 gene (D85Y) is not associated with prostate cancer risk. *Cancer Epidemiol Biomarkers Prev* **11**:497–498.

Guha C, Parashar B, Deb NJ, Garg M, Gorla GR, Singh A, Roy-Chowdhury N, Vikram B, and Roy-Chowdhury J (2002) Long-term normalization of serum bilirubin levels by massive repopulation of Gunn rat liver by normal hepatocytes, transplanted after preparative hepatic irradiation and partial hepatectomy. *Hepatology* **36**:354–362.

Guha C, Sharma A, Gupta S, Alfieri A, Gorla GR, Gagandeep S, Sokhi R, Roy-Chowdhury N, Tanaka KE, Vikram B, and Roy-Chowdhury J (1999) Amelioration of radiation-induced liver damage in partially hepatectomized rats by hepatocyte transplantation. *Cancer Res* **59**:5871–5874.

Guillemette C, De Vivo I, Hankinson SE, Haiman CA, Spiegelman D, Housman DE, and Hunter DJ (2001) Association of genetic polymorphisms in UGT1A1 with breast cancer and plasma hormone levels. *Cancer Epidemiol Biomarkers Prev* **10**:711–714.

Guillemette C, Levesque E, Beaulieu M, Turgeon D, Hum DW, and Belanger A (1997) Differential regulation of two uridine diphospho-glucuronosyltransferases, UGT2B15 and UGT2B17, in human prostate LNCaP cells. *Endocrinology* **138**:2998–3005.

Guillemette C, Millikan R, Newman B, and Housman DE (2000a) Genetic polymorphisms in UGT1A1 and association with breast cancer among African Americans. *Cancer Res* **60**:950–956.

Guillemette C, Ritter JK, Auyeung DJ, Kessler FK, and Housman DE (2000b) Structural heterogeneity at the UDP-glucuronosyltransferase 1A locus: functional consequences of three novel missense mutations in the human UGT1A7 gene. *Pharmacogenetics* **10**:629–644.

Guillemette CG (2003) Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *Pharmacogenom J* **3**:136–158.

Gupta S, Aragona E, Vemuru RP, Bhargava KK, Burk RD, and Chowdhury JR (1991) Permanent engraftment and function of hepatocytes delivered to the liver: implications for gene therapy and liver repopulation. *Hepatology* **14**:144–149.

Haiman CA, Hankinson SE, De Vivo I, Guillemette C, Ishibe N, Hunter DJ, and Byrne C (2003) Polymorphisms in steroid hormone pathway genes and mammographic density. *Breast Cancer Res Treat* **77**:27–36.

Hall D, Ybazeta G, Destro-Bisol G, Petzl-Erler ML, and Di Rienzo A (1999) Variability at the uridine diphosphate glucuronosyltransferase 1A1 promoter in human populations and primates. *Pharmacogenetics* **9**:591–599.

Hansen AJ, Lee YH, Gonzalez FJ, and Mackenzie PI (1997) HNF1 alpha activates the rat UDP glucuronosyltransferase UGT2B1 gene promoter. *DNA Cell Biol* **16**:207–214.

Hansen AJ, Lee YH, Sterneck E, Gonzalez FJ, and Mackenzie PI (1998) C/EBPalpha is a regulator of the UDP glucuronosyltransferase UGT2B1 gene. *Mol Pharmacol* **53**:1027–1033.

Haque SJ, Petersen DD, Nebert DW, and Mackenzie PI (1991) Isolation, sequence and developmental expression of rat UGT2B2: the gene encoding a constitutive UDP glucuronosyltransferase that metabolizes etiocholanolone and androsterone. *DNA Cell Biol* **10**:515–524.

Hu Z and Wells PG (1992) In vitro and in vivo biotransformation and covalent binding of benzo(a)pyrene in rats with a genetic deficiency in bilirubin UDP-glucuronosyltransferase. *J Pharmacol Exp Ther* **263**:334–342.

Hu Z and Wells PG (1993) Human interindividual variation in lymphocyte UDP-glucuronosyltransferases as a determinant of benzo[a]pyrene covalent binding and cytotoxicity. *ISSX Proc* 4:241.

Hu Z and Wells PG (1994) Modulation of benzo[a]pyrene bioactivation and cytotoxicity by glucuronidation in lymphocytes and hepatic microsomes from rats with a hereditary deficiency in bilirubin UDP-glucuronosyltransferase. *Toxicol Appl Pharmacol* **127**:306–313.

Huang YH, Galijatovic A, Nguyen N, Geske D, Beaton D, Green J, Green M, Peters WH, and Tukey RH (2002) Identification and functional characterization of UDP-glucuronosyltransferases UGT1A8*1, UGT1A8*2 and UGT1A8*3. *Pharmacogenetics* **12**:287–297.

Ikushiro S, Emi Y, and Iyanagi T (1995) Identification and analysis of drug-responsive expression of UDP-glucuronosyltransferase family 1 (UGT1) isozyme in rat hepatic microsomes using anti-peptide antibodies. *Arch Biochem Biophys* **324**:267–272.

Ilan Y, Attavar P, Takahashi M, Davidson A, Horwitz M, Guida J, Roy Chowdhury N, and Roy Chowdhury J (1996) Induction of central tolerance by intrathymic inoculation of adenoviral antigens into the host thymus permits long-term gene therapy in Gunn rats. *J Clin Investig* **98**:2640–2647.

Ilan Y, Prakash R, Davidson A, Jona V, Droguett G, Horwitz MS, Roy Chowdhury N, and Roy Chowdhury J (1997) Oral tolerization to adenoviral antigens permits long-term gene expression using recombinant adenoviral vectors. *J Clin Investig* **99**:1098–1106.

Ishii Y, Hansen AJ, and Mackenzie PI (2000) Octamer transcription factor-1 enhances hepatic nuclear factor-1 alpha-mediated activation of the human UDP glucuronosyltransferase 2B7 promoter. *Mol Pharmacol* **57**:940–947.

Iyanagi T, Emi Y, and Ikushiro S (1998) Biochemical and molecular aspects of genetic disorders of bilirubin metabolism. *Biochim Biophys Acta* **1407**:173–174.

Iyer L, King CD, Whitington PF, Green MD, Roy SK, Tephly TR, Coffman BL, and Ratain MJ (1998) Genetic predisposition to the metabolism of irinotecan (CPT-11). Role of uridine diphosphate glucuronosyltransferase isoform 1A1 in the glucuronidation of its active metabolite (SN-38) in human liver microsomes. *J Clin Investig* **101**:847–854.

Izsvák Z, Ivics Z, and Plasterk RH (2000) Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *J Mol Biol* **302**:93–102.

Jinno H, Saeki M, Saito Y, Tanaka-Kagawa T, Hanioka N, Sai K, Kaniwa N, Ando M, Shirao K, Minami H, et al. (2003a) Functional characterization of human UDP- glucuronosyltransferase (UGT) 1A9 variant, D256N, found in Japanese cancer patients. *J Pharmacol Exp Ther* **306**:688–693.

Jinno H, Saeki M, Tanaka-Kagawa T, Hanioka N, Saito Y, Ozawa S, Ando M, Shirao K, Minami H, Ohtsu A, et al. (2003b) Functional characterization of wild-type and variant (T202I and M59I) human UDP-glucuronosyltransferase 1A10. *Drug Metab Dispos* **31**:528–532.

Kessler FK, Kessler MR, Auyeung DJ, and Ritter JK (2002) Glucuronidation of acetaminophen catalyzed by multiple rat phenol UDP-glucuronosyltransferases. *Drug Metab Dispos* **30**:324–330.

Kessler FK and Ritter JK (1997) Induction of a rat liver benzo[a]pyrene-trans-7,8-dihydrodiol glucuronidating activity by oltipraz and beta-naphthoflavone. *Carcinogenesis* **18**:107–114.

Kim PM, DeBoni U, and Wells PG (1997a) Peroxidase-dependent bioactivation and oxidation of DNA and protein in benzo[a]pyrene-initiated micronucleus formation. *Free Radic Biol Med* **23**:579–596.

Kim PM and Wells PG (1996a) Genoprotection by UDP-glucuronosyltransferases in peroxidase-dependent, reactive oxygen species-mediated micronucleus initiation by the carcinogens 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) and benzo[a]pyrene. *Cancer Res* **56**: 1526–1532.

Kim PM and Wells PG (1996b) Phenytoin-initiated hydroxyl radical formation: characterization by enhanced salicylate hydroxylation. *Mol Pharmacol* **49**:172–181.

Kim PM and Wells PG (1998) Phenytoin embryotoxicity: protection by UDP-glucuronosyltransferases. *Toxicol Sci* **42(1-S)**: 261 (No. 1288).

Kim PM, Winn LM, Parman T, and Wells PG (1997b) UDP-glucuronosyltransferase-mediated protection against in vitro DNA oxidation and micronucleus formation initiated by phenytoin and its embryotoxic metabolite 5-(*p*-hydroxyphenyl)-5-phenylhydantoin (HPPH). *J Pharmacol Exp Ther* **280**:200–209.

Kobayashi T, Yokota H, Ohgiya S, Iwano H, and Yuasa A (1998) UDP-glucuronosyltransferase UGT1A7 induced in rat small intestinal mucosa by oral administration of 2-naphthoflavone. *Eur J Biochem* **258**:948–955.

Kren B, Parashar B, Bandopadhyay P, Roy Chowdhury N, Roy Chowdhury J, and Steer CJ (1999) Correction of the UDP-glucuronosyltransferase gene defect in the Gunn rat model of Crigler-Najjar syndrome type I. *Proc Natl Acad Sci USA* **96**:10349–10354.

Lampe JW, Bigler J, Bush AC, and Potter JD (2000). Prevalence of polymorphisms in the human UDP-glucuronosyltransferase 2B family: UGT2B4(D458E), UGT2B7(H268Y), and UGT2B15(D85Y). *Cancer Epidemiol Biomarkers Prev* **9**:329–333.

Lampe JW, Bigler J, Horner NK, and Potter JD (1999) UDP-glucuronosyltransferase (UGT1A1*28 and UGT1A6*2) polymorphisms in Caucasians and Asians: relationships to serum bilirubin concentrations. *Pharmacogenetics* **9**:341–349.

Lee YH, Sauer B, Johnson PF, and Gonzalez FJ (1997) Disruption of the C/Ebp-alpha gene in adult mouse liver. *Mol Cell Biol* **17**:6014–6022.

Levesque E, Beaulieu M, Green MD, Tephly TR, Belanger A, and Hum DW (1997) Isolation and characterization of UGT2B15(Y85): a UDP-glucuronosyltransferase encoded by a polymorphic gene. *Pharmacogenetics* **7**:317–325.

Levesque E, Beaulieu M, Hum DW, and Belanger A (1999) Characterization and substrate specificity of UGT2B4 (E458): a UDP-glucuronosyltransferase encoded by a polymorphic gene. *Pharmacogenetics* **9**:207–216.

Levesque EE, Turgeon D, Carrier JS, Montminy V, Beaulieu M, and Belanger A (2001) Isolation and characterization of the UGT2B28 cDNA encoding a novel human steroid conjugating UDP-glucuronosyltransferase. *Biochemistry* **40**:3869–3881.

Lichtsteiner S, Wuarin J, and Schibler U (1987) The interplay of DNA-binding proteins on the promoter of the mouse albumin gene. *Cell* **51**:963–973.

Mackenzie PI, Owens IS, Burchell B, Bock KW, Bairoch A, Bélanger A, Fournel-Gigleux S, Green M, Hum DW, Iyanagi T, et al. (1997) The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics* **7**:255–269.

Mackenzie PI and Rodbourn L (1990) Organization of the rat UDP-glucuronosyltransferase, UDPGTr-2, gene and characterization of its promoter. *J Biol Chem* **265**:11328–11332.

MacLeod SL, Nowell S, Plaxco J, and Lang NP (2000) An allele-specific polymerase chain reaction method for the determination of the D85Y polymorphism in the human UDP-glucuronosyltransferase 2B15 gene in a case-control study of prostate cancer. *Ann Surg Oncol* **7**:777–782.

Maruo Y, Nishizawa K, Sato H, Doida Y, and Shimada M (1999) Association of neonatal hyperbilirubinemia with bilirubin UDP-glucuronosyltransferase polymorphism. *Pediatrics* **103**:1224–1227.

Mojarrabi B and Mackenzie PI (1998) Characterization of two UDP glucuronosyltransferases that are predominantly expressed in human colon. *Biochem Biophys Res Commun* **247**:704–709.

Monaghan G, Burchell B, and Boxer M (1997) Structure of the human Ugt2b4 gene encoding a bile acid UDP-glucuronosyltransferase. *Mamm Genome* **8**:692–694.

Nakai H, Yant SR, Storm TA, Fuess S, Meuse L, and Kay MA (2001) Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo. *J Virol* **75**:6969–6976.

Owens IS and Ritter JK (1992) The novel bilirubin/phenol UDP-glucuronosyltransferase UGT1 gene locus: implications for multiple nonhemolytic familial hyperbilirubinemia phenotypes. *Pharmacogenetics* **2**:93–108.

Radominska-Pandya A, Czernik PJ, Little JM, Battaglia E, and Mackenzie PI (1999) Structural and functional studies of UDP-glucuronosyltransferases. *Drug Metab Rev* **31**:817–899.

Riedy M, Wang JY, Miller AP, Buckler A, Hall J, and Guida M (2000) Genomic organization of the UGT2b gene cluster on human chromosome 4q13. *Pharmacogenetics* **10**:251–260.

Ritter JK, Chen F, Sheen YY, Tran HM, Kimura S, Yeatman MT, and Owens IS (1992) A novel complex locus UGT1 encodes human bilirubin, phenol and other UDP-glucuronosyltransferase isozymes with identical carboxyl termini. *J Biol Chem* **267**:3257–3261.

Ritter JK, Kessler FK, Thompson MT, Grove AD, Auyeung DJ, and Fisher RA (1999) Expression and inducibility of the human bilirubin UDP-glucuronosyltransferase UGT1A1 in liver and cultured primary hepatocytes: evidence for both genetic and environmental influences. *Hepatology* **30**:476–484.

Roy Chowdhury J, Grossman M, Gupta S, Roy Chowdhury N, Baker JR Jr, and Wilson JM (1991) Long-term improvement of hypercholesterolemia after ex vivo gene therapy in LDL-receptor deficient rabbits. *Science (Wash DC)* **254**:1802–1805.

Roy Chowdhury N, Hays RM, Bommineni VR, Franki N, Roy Chowdhury J, Wu CH, and Wu GY (1996) Microtubular distribution prolongs the expression of human bilirubin-uridinediphosphoglucuronate-glucuronosyltransferase-1 gene transferred into Gunn rat livers. *J Biol Chem* **271**:2341–2346.

Roy Chowdhury J, Roy Chowdhury N, Strom SC, Kaufman SS, Horslen S, and Fox IJ (1998) Hepatocyte transplantation in humans: gene therapy and more. *Pediatrics* **102**:647–648.

Roy Chowdhury J, Wolkoff AW, Roy Chowdhury N, and Arias IM (2001) Hereditary jaundice and disorders of bilirubin metabolism, in *The Metabolic and Molecular Basis of Inherited Disease*, 8th ed (Scriver CR, Boudet AL, Sly W, Valle D, Childs B, Kinzler K, and Vogelstein B, eds) pp 3063–3101, McGraw-Hill, New York.

Sauter BV, Parashar B, Roy Chowdhury N, Kadakol A, Ilan Y, Singh H, Milano J, Strayer DS, and Roy Chowdhury J (2000) Gene transfer to the liver using a replication-deficient recombinant SV40 vector results in long-term amelioration of jaundice in Gunn rats. *Gastroenterology* **119**:1348–1357.

Senafi SB, Clarke DJ, and Burchell B (1994) Investigation of the substrate specificity of a cloned expressed human bilirubin UDP-glucuronosyltransferase: UDP-sugar specificity and involvement in steroid and xenobiotic glucuronidation. *Biochem J* **303**:233–240.

Strassburg CP, Nguyen N, Manns MP, and Tukey RH (1998) Polymorphic expression of the UDP-glucuronosyltransferase UGT1A gene locus in human gastric epithelium. *Mol Pharmacol* **54**:647–654.

Strassburg CP, Vogel A, Kneip S, Tukey RH, and Manns MP (2002) Polymorphisms of the human UDP-glucuronosyltransferase (UGT) 1A7 gene in colorectal cancer. *Gut* **50**:851–856.

Strayer DS, Branco F, Zern MA, Yam P, Calarota SA, Nichols CN, Zaia JA, Rossi J, Li H, Parashar B, et al. (2002) Durability of transgene expression and vector integration: recombinant SV40-derived gene therapy vectors. *Mol Ther* **6**:227–237.

Tada K, Roy Chowdhury N, Prasad VR, Kim B-H, Kalapudi M, Fox IJ, Duijvandijk P, Bosma PJ, and Roy Chowdhury J (1998) Long-term amelioration of bilirubin glucuronidation defect in Gunn rats by transplanting genetically modified immortalized autologous hepatocytes. *Cell Transplant* **7**:607–616.

Takahashi M, Ilan Y, Roy Chowdhury N, Guida J, Horwitz MS, and Roy Chowdhury J (1996) Long-term correction of bilirubin UDP-glucuronosyltransferase deficiency in Gunn rats by administration of a recombinant adenovirus during the neonatal period. *J Biol Chem* **271**: 26536–26542.

Thummala NR, Ghosh SS, Lee SW, Reddy B, Davidson A, Horwitz MS, Roy Chowdhury J, and Roy Chowdhury N (2002) A non-immunogenic adenoviral vector, coexpressing CTLA4Ig and bilirubin-uridinediphosphoglucuronateglucuronosyltransferase permits long-term, repeatable transgene expression in the Gunn rat model of Crigler-Najjar syndrome. *Gene Ther* **9**:981–990.

Tukey RH and Strassburg CP (2000) Human UDP-glucuronosyltransferases: metabolism, expression and disease. *Annu Rev Pharmacol Toxicol* **40**:581–616.

Turgeon D, Carrier JS, Levesque E, Beatty BG, Belanger A, and Hum DW (2000) Isolation and characterization of the human UGT2B15 gene, localized within a cluster of UGT2B genes and pseudogenes on chromosome 4. *J Mol Biol* **295**:489–504.

Ullrich D, Sieg A, Blume R, Bock KW, Schroter W, and Bircher J (1987) Normal pathways for glucuronidation, sulphation and oxidation of paracetamol in Gilbert's syndrome. *Eur J Clin Investig* **17**:237–240.

Vienneau DS, DeBoni U, and Wells PG (1995) Potential genoprotective role for UDP-glucuronosyltransferases (UGTs) in chemical carcinogenesis: initiation of micronuclei by benzo[a]pyrene and benzo[e]pyrene in UGT-deficient cultured rat skin fibroblasts. *Cancer Res* **55**:1045–1051.

Villeneuve L, Girard H, Fortier LC, and Guillemette C (2003) Novel functional polymorphisms in the UGT1A7 and UGT1A9 glucuronidating enzymes in Caucasian and African-American subjects and their impact on the metabolism of 7-ethyl-10-hydroxycamptothecin and flavopiridol anticancer drugs. *J Pharmacol Exp Ther* **307**:117–128.

Vogel A, Kneip S, Barut A, Ehmer U, Tukey RH, Manns MP, and Strassburg CP (2001) Genetic link of hepatocellular carcinoma with polymorphisms of the UDP-glucuronosyltransferase UGT1A7 gene. *Gastroenterology* **121**:1136–1144.

Wells PG, Kim PM, Nicol CJ, Parman T, and Winn LM (1997) Chapter 17: Reactive intermediates, in *Handbook of Experimental Pharmacology, Part I: Drug Toxicity in Embryonic Development* (Kavlock RJ and Daston GP eds.), vol 124, pp 453–518, Springer-Verlag, Heidelberg.

Wells PG, Obilo FC, and de Morais SMF (1989) Benzo(a)pyrene embryopathy in rats genetically deficient in bilirubin UDP-glucuronyl transferase. *FASEB J* **3**:A1025.

Wells PG and Winn LM (1996) Biochemical toxicology of chemical teratogenesis. *Crit Rev Biochem Mol Biol* **31**:1–40.

Woolley AT, Guillemette C, Cheung CL, Housman DE, and Lieber CM (2000) Direct haplotyping of kilobase-size DNA using carbon nanotube probes. *Nat Biotechnol* **18**:760–763.

Zheng Z, Park JY, Guillemette C, Schantz SP, and Lazarus P (2001) Tobacco carcinogen-detoxifying enzyme UGT1A7 and its association with orolaryngeal cancer risk. *J Natl Cancer Inst* **93**:1411–1418.

Exhibit A
with Response dated 4/26/04
In USSN: 09/980,729

1: NP_001066. UDP glycosyltrans...[gi:4507817] BLink, Domains, Links

```
LOCUS       NP_001066                528 aa            linear   PRI 23-DEC-2003
DEFINITION  UDP glycosyltransferase 2 family, polypeptide B10 [Homo sapiens].
ACCESSION   NP_001066
VERSION     NP_001066.1  GI:4507817
DBSOURCE    REFSEQ: accession NM_001075.2
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Jin,C.J., Miners,J.O., Lillywhite,K.J. and Mackenzie,P.I.
  TITLE     cDNA cloning and expression of two new members of the human liver
            UDP-glucuronosyltransferase 2B subfamily
  JOURNAL   Biochem. Biophys. Res. Commun. 194 (1), 496-503 (1993)
   PUBMED   8333863
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from X63359.1.
FEATURES             Location/Qualifiers
     source          1..528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4q13.3"
     Protein         1..528
                     /product="UDP glycosyltransferase 2 family, polypeptide
                     B10"
                     /EC_number="2.4.1.17"
     Region          23..526
                     /region_name="UDP-glucoronosyl and UDP-glucosyl
                     transferase"
                     /note="UDPGT"
                     /db_xref="CDD:22944"
     variation       283
                     /replace="P"
                     /replace="A"
                     /db_xref="dbSNP:1976666"
     variation       382
                     /replace="A"
                     /replace="T"
                     /db_xref="dbSNP:4095564"
     CDS             1..528
                     /gene="UGT2B10"
                     /coded_by="NM_001075.2:11..1597"
                     /note="go_component: microsome [goid 0005792] [evidence
                     IEA];
                     go_component: integral to membrane [goid 0016021]
                     [evidence IEA];
                     go_function: UDP-glucuronosyltransferase [goid 0003981]
                     [evidence E] [pmid 8333863];
                     go_function: glucuronosyltransferase activity [goid
                     0015020] [evidence IEA];
                     go_process: lipid metabolism [goid 0006629] [evidence TAS]
                     [pmid 8333863]"
                     /db_xref="GeneID:7365"
                     /db_xref="LocusID:7365"
                     /db_xref="MIM:600070"
ORIGIN
        1 malkwttvll iqlsfyfssg scgkvlvwaa eyslwmnmkt ilkelvqrgh evtvlassas
       61 ilfdpndsst lklevyptsl tktefeniim qlvkrlseiq kdtfwlpfsq eqeilwaind
      121 iirnfckdvv snkklmkklq esrfdivfad aylpcgella elfnipfvys hsfspgysfe
      181 rhsggfifpp syvpvvmskl sdqmtfmerv knmlyvlyfd fwfqifnmkk wdqfysevlg
```

```
      241 rpttlsetmr kadiwlmrns wnfkfphpfl pnvdfvgglh ckpakplpke meefvqssge
      301 ngvvvfslgs mvsnmteera nviatalaki pqkvlwrfdg nkpdalglnt rlykwipqnd
      361 llghpktraf ithggangiy eaiyhgipmv giplffdqpd niahmkakga avrvdfntms
      421 stdllnalkt vindpsyken imklsriqhd qpvkpldrav fwiefvmrhk gakhlrvaah
      481 nltwfqyhsl dvigfllacv atvlfiitkc clfcfwkfar kgkkgkrd
//
```

SeqServer®
biology in silico

ClustalW Results

Sequences | Help

Retrieval | BLAST2 | FASTA | ClustalW | GCG Assembly | Phrap | Translation

 SeqServer Version 4.6 Jan 2002

- 2912330CD1
- g4507817

CLUSTAL W (1.7) Multiple Sequence Alignments

```
Sequence format is Pearson
Sequence 1: 2912330CD1      529 aa
Sequence 2: g4507817        528 aa
Start of Pairwise alignments
Aligning...
Sequences (1:2) Aligned. Score:  80
Start of Multiple Alignment
There are 1 groups
Aligning...
Group 1: Sequences:   2      Score:6470
Alignment Score 2761
CLUSTAL-Alignment file created  [baaubaiwy.aln]
CLUSTAL W (1.7) multiple sequence alignment


2912330CD1      MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA
g4507817        MALKWT-TVLLIQLSFYFSSGSCGKVLVWAAEYSLWMNMKTILKELVQRGHEVTVLASSA
                *::*** ::****** ************.:*:* ***:****.****************

2912330CD1      SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPKDTFWSYFSQVQEIMWTFN
g4507817        SILFDPNDSSTLKLEVYPTSLTKTEFENIIMQLVKRLSEIQKDTFWLPFSQEQEILWAIN
                ** ****..****:****.********:** ***** :*: *****  *** ***:*::*

2912330CD1      DILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYAI
g4507817        DIIRNFCKDVVSNKKLMKKLQESRFDIVFADAYLPCGELLAELFNIPFVYSHSFSPGYSF
                **:*:****:****************:*:*** :* *******::******  *****::

2912330CD1      EKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEVL
g4507817        ERHSGGFIFPPSYVPVVMSKLSDQMTFMERVKNMLYVLYFDFWFQIFNMKKWDQFYSEVL
                *:****::***********:*******:******:*****:******:************

2912330CD1      GRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSSG
g4507817        GRPTTLSETMRKADIWLMRNSWNFKFPHPFLPNVDFVGGLHCKPAKPLPKEMEEFVQSSG
                ********** ******:** *:*:****:****:************************

2912330CD1      ENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQN
g4507817        ENGVVVFSLGSMVSNMTEERANVIATALAKIPQKVLWRFDGNKPDALGLNTRLYKWIPQN
```

```
                ********************* :********:********************:**************

2912330CD1      DLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTM
g4507817        DLLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFFDQPDNIAHMKAKGAAVRVDFNTM
                ********:******* **********:****:*:* ** *************.::*:**

2912330CD1      TSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAA
g4507817        SSTDLLNALKTVINDPSYKENIMKLSRIQHDQPVKPLDRAVFWIEFVMRHKGAKHLRVAA
                :* ***.**:***.*.***** *:****:**************************** **

2912330CD1      HDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE
g4507817        HNLTWFQYHSLDVIGFLLACVATVLFIITKCCLFCFWKFARKGKKGKRD
                *:*****::*:*******:****.:*::*** **.  ** :. *  **:
```

Submit sequences to: BLAST2 Submit Reset

IncyteGenomics

1: P36537. UDP-glucuronosylt...[gi:549155] BLink, Domains, Links

```
LOCUS       P36537                  528 aa            linear   PRI 15-JUN-2002
DEFINITION  UDP-glucuronosyltransferase 2B10 precursor, microsomal (UDPGT).
ACCESSION   P36537
VERSION     P36537  GI:549155
DBSOURCE    swissprot: locus UDBA_HUMAN, accession P36537;
            class: standard.
            created: Jun 1, 1994.
            sequence updated: Jun 1, 1994.
            annotation updated: Jun 15, 2002.
            xrefs: gi: 516149, gi: 516150, gi: 484384
            xrefs (non-sequence databases): MIM 600070, InterProIPR002213,
            PfamPF00201, PROSITEPS00375
KEYWORDS    Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
            Signal; Multigene family; Microsome.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 528)
  AUTHORS   Jin,C.J., Miners,J.O., Lillywhite,K.J. and Mackenzie,P.I.
  TITLE     cDNA cloning and expression of two new members of the human liver
            UDP-glucuronosyltransferase 2B subfamily
  JOURNAL   Biochem. Biophys. Res. Commun. 194 (1), 496-503 (1993)
  MEDLINE   93326164
   PUBMED   8333863
  REMARK    SEQUENCE FROM N.A.
            TISSUE=Liver
COMMENT     ---------------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------------.
            [FUNCTION] UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
            SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
            ENDOGENOUS COMPOUNDS.
            [CATALYTIC ACTIVITY] UDP-glucuronate + acceptor = UDP + acceptor
            beta-D-glucuronoside.
            [SUBCELLULAR LOCATION] Microsomal.
            [SIMILARITY] BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
FEATURES             Location/Qualifiers
     source          1..528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
     gene            1..528
                     /gene="UGT2B10"
     Protein         1..528
                     /gene="UGT2B10"
                     /product="UDP-glucuronosyltransferase 2B10 precursor,
                     microsomal"
                     /EC_number="2.4.1.17"
     Region          1..23
                     /gene="UGT2B10"
                     /region_name="Signal"
                     /note="BY SIMILARITY."
     Region          24..528
                     /gene="UGT2B10"
                     /region_name="Mature chain"
                     /note="UDP-GLUCURONOSYLTRANSFERASE 2B10."
     Site            66
                     /gene="UGT2B10"
```

```
                     /site_type="glycosylation"
                     /note="N-LINKED (GLCNAC...) (POTENTIAL)."
     Site            314
                     /gene="UGT2B10"
                     /site_type="glycosylation"
                     /note="N-LINKED (GLCNAC...) (POTENTIAL)."
     Site            481
                     /gene="UGT2B10"
                     /site_type="glycosylation"
                     /note="N-LINKED (GLCNAC...) (POTENTIAL)."
     Region          492..512
                     /gene="UGT2B10"
                     /region_name="Transmembrane region"
                     /note="POTENTIAL."
ORIGIN
        1 malkwttvll iqlsfyfssg scgkvlvwaa eyslwmnmkt ilkelvqrgh evtvlassas
       61 ilfdpndsst lklevyptsl tktefeniim qlvkrlseiq kdtfwlpfsq eqeilwaind
      121 iirnfckdvv snkklmkklq esrfdivfad aylpcgella elfnipfvys hsfspgysfe
      181 rhsggfifpp syvpvvmskl sdqmtfmerv knmlyvlyfd fwfqifnmkk wdqfysevlg
      241 rpttlsetmr kadiwlmrns wnfkfphpfl pnvdfvgglh ckpakplpke meefvqssge
      301 ngvvvfslgs mvsnmteera nviatalaki pqkvlwrfdg nkpdalglnt rlykwipqnd
      361 llghpktraf ithggangiy eaiyhgipmv giplffdqpd niahmkakga avrvdfntms
      421 stdllnalkt vindpsyken imklsriqhd qpvkpldrav fwiefvmrhk gakhlrvaah
      481 nltwfqyhsl dvigfllacv atvlfiitkc clfcfwkfar kgkkgkrd
//
```


Jan 29 2004 15:38:25

SeqServer®
biology *in silico*

ClustalW Results

Sequences | Help

Retrieval | BLAST2 | FASTA | ClustalW | GCG Assembly | Phrap | Translation

 SeqServer Version 4.6 Jan 2002

- 2912330CD1
- g549155

CLUSTAL W (1.7) Multiple Sequence Alignments

```
Sequence format is Pearson
Sequence 1: 2912330CD1      529 aa
Sequence 2: g549155         528 aa
Start of Pairwise alignments
Aligning...
Sequences (1:2) Aligned. Score:  80
Start of Multiple Alignment
There are 1 groups
Aligning...
Group 1: Sequences:   2      Score:6470
Alignment Score 2761
CLUSTAL-Alignment file created  [baaOSayOy.aln]
CLUSTAL W (1.7) multiple sequence alignment


2912330CD1      MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA
g549155         MALKWT-TVLLIQLSFYFSSGSCGKVLVWAAEYSLWMNMKTILKELVQRGHEVTVLASSA
                *::*** ::****** ************.:*:* ***:****.****************

2912330CD1      SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPKDTFWSYFSQVQEIMWTFN
g549155         SILFDPNDSSTLKLEVYPTSLTKTEFENIIMQLVKRLSEIQKDTFWLPFSQEQEILWAIN
                ** ****..****:****.********:** ***** :*: *****  *** ***:*::*

2912330CD1      DILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYAI
g549155         DIIRNFCKDVVSNKKLMKKLQESRFDIVFADAYLPCGELLAELFNIPFVYSHSFSPGYSF
                **:*:****:****************:*:*** :* *******::******  *****::

2912330CD1      EKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEVL
g549155         ERHSGGFIFPPSYVPVVMSKLSDQMTFMERVKNMLYVLYFDFWFQIFNMKKWDQFYSEVL
                *:****::***********:*******:******:*****:******:************

2912330CD1      GRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSSG
g549155         GRPTTLSETMRKADIWLMRNSWNFKFPHPFLPNVDFVGGLHCKPAKPLPKEMEEFVQSSG
                ********** ******:** *:*:****:****:************************

2912330CD1      ENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQN
g549155         ENGVVVFSLGSMVSNMTEERANVIATALAKIPQKVLWRFDGNKPDALGLNTRLYKWIPQN
```

```
               *************** :********:********************:**************

2912330CD1     DLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTM
g549155        DLLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFFDQPDNIAHMKAKGAAVRVDFNTM
               ********:******* **********:****:*:* ** *************.::*:**

2912330CD1     TSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAA
g549155        SSTDLLNALKTVINDPSYKENIMKLSRIQHDQPVKPLDRAVFWIEFVMRHKGAKHLRVAA
               :* ***.**:***.*.***** *:****:**************************** **

2912330CD1     HDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE
g549155        HNLTWFQYHSLDVIGFLLACVATVLFIITKCCLFCFWKFARKGKKGKRD
               *:*****::*:*******:****.:*::*** **.  ** :. *  **:
```

Submit sequences to: BLAST2 Submit Reset

IncyteGenomics

1: NP_001064. UDP glycosyltrans...[gi:4507823] BLink, Domains, Links

```
LOCUS       NP_001064                529 aa            linear   PRI 23-DEC-2003
DEFINITION  UDP glycosyltransferase 2 family, polypeptide B11 [Homo sapiens].
ACCESSION   NP_001064
VERSION     NP_001064.1  GI:4507823
DBSOURCE    REFSEQ: accession NM_001073.1
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 529)
  AUTHORS   Beaulieu,M., Levesque,E., Hum,D.W. and Belanger,A.
  TITLE     Isolation and characterization of a human orphan
            UDP-glucuronosyltransferase, UGT2B11
  JOURNAL   Biochem. Biophys. Res. Commun. 248 (1), 44-50 (1998)
   PUBMED   9675083
REFERENCE   2  (residues 1 to 529)
  AUTHORS   Jin,C.J., Miners,J.O., Lillywhite,K.J. and Mackenzie,P.I.
  TITLE     cDNA cloning and expression of two new members of the human liver
            UDP-glucuronosyltransferase 2B subfamily
  JOURNAL   Biochem. Biophys. Res. Commun. 194 (1), 496-503 (1993)
   PUBMED   8333863
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from AF016492.1.
FEATURES             Location/Qualifiers
     source          1..529
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4q13.3"
     Protein         1..529
                     /product="UDP glycosyltransferase 2 family, polypeptide
                     B11"
     Region          24..527
                     /region_name="UDP-glucoronosyl and UDP-glucosyl
                     transferase"
                     /note="UDPGT"
                     /db_xref="CDD:22944"
     variation       70
                     /replace="P"
                     /replace="S"
                     /db_xref="dbSNP:7697482"
     variation       101
                     /replace="Q"
                     /replace="R"
                     /db_xref="dbSNP:7688262"
     variation       156
                     /replace="R"
                     /replace="C"
                     /db_xref="dbSNP:7697037"
     variation       289
                     /replace="P"
                     /replace="L"
                     /db_xref="dbSNP:3890590"
     CDS             1..529
                     /gene="UGT2B11"
                     /coded_by="NM_001073.1:10..1599"
                     /note="go_component: microsome [goid 0005792] [evidence
                     IEA];
                     go_component: integral to membrane [goid 0016021]
                     [evidence IEA];
                     go_function: UDP-glucuronosyltransferase [goid 0003981]
```

```
                     [evidence E] [pmid 8333863];
                     go_function: glucuronosyltransferase activity [goid
                     0015020] [evidence IEA];
                     go_process: estrogen metabolism [goid 0008210] [evidence
                     TAS] [pmid 8333863];
                     go_process: xenobiotic metabolism [goid 0006805] [evidence
                     TAS] [pmid 8333863]"
                     /db_xref="GeneID:10720"
                     /db_xref="LocusID:10720"
                     /db_xref="MIM:603064"
ORIGIN
        1 mtlkwtsvll lihlscyfss gscgkvlvwa aeyshwmnmk tilkelvqrg hevtvlassa
       61 silfdpndas tlkfevypts ltktefenii mqqvkrwsdi rkdsfwlyfs qeqeilwely
      121 difrnfckdv vsnkkvmkkl qesrfdivfa davfpcgell aallnirfvy slrftpgyti
      181 erhsgglifp psyipivmsk lsdqmtfmer vknmiyvlyf dfwfqmsdmk kwdqfysevl
      241 grpttlfetm gkadiwlmrn swsfqfphpf lpnvdfvggf hckpakplpk emeefvqssg
      301 engvvvfslg svisnmtaer anviatalak ipqkvlwrfd gnkpdalgln trlykwipqn
      361 dllghpktra fithggangi yeaiyhgipm vgiplffdqp dniahmkakg aavrldfntm
      421 sstdllnalk tvindplyke nimklsriqh dqpvkpldra vfwiefvmph kgakhlrvaa
      481 hdltwfqyhs ldvigfllac vatvifiitk fclfcfwkfa rkgkkgkrd
//
```

SeqServer®
biology *in silico*

ClustalW Results

Sequences | Help

Retrieval | BLAST2 | FASTA | ClustalW | GCG Assembly | Phrap | Translation

- 2912330CD1
- g4507823

CLUSTAL W (1.7) Multiple Sequence Alignments

```
Sequence format is Pearson
Sequence 1: 2912330CD1      529 aa
Sequence 2: g4507823        529 aa
Start of Pairwise alignments
Aligning...
Sequences (1:2) Aligned. Score:  80
Start of Multiple Alignment
There are 1 groups
Aligning...
Group 1: Sequences:   2      Score:6500
Alignment Score 2777
CLUSTAL-Alignment file created  [baadWaOez.aln]
CLUSTAL W (1.7) multiple sequence alignment


2912330CD1      MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA
g4507823        MTLKWTSVLLLIHLSCYFSSGSCGKVLVWAAEYSHWMNMKTILKELVQRGHEVTVLASSA
                *::****.****:****************.:*:*****:****.****************

2912330CD1      SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPKDTFWSYFSQVQEIMWTFN
g4507823        SILFDPNDASTLKFEVYPTSLTKTEFENIIMQQVKRWSDIRKDSFWLYFSQEQEILWELY
                ** ****..*********.********:** * ****::: **:** **** ***:* :

2912330CD1      DILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYAI
g4507823        DIFRNFCKDVVSNKKVMKKLQESRFDIVFADAVFPCGELLAALLNIRFVYSLRFTPGYTI
                **:*:****:*****:**********:*:****** ***** **:* *******:***:*

2912330CD1      EKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEVL
g4507823        ERHSGGLIFPPSYIPIVMSKLSDQMTFMERVKNMIYVLYFDFWFQMSDMKKWDQFYSEVL
                *:*****:*****:*:***:*******:***********:****: ************

2912330CD1      GRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSSG
g4507823        GRPTTLFETMGKADIWLMRNSWSFQFPHPFLPNVDFVGGFHCKPAKPLPKEMEEFVQSSG
                ****** ***.******:** *.******:****:****:********************

2912330CD1      ENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQN
g4507823        ENGVVVFSLGSVISNMTAERANVIATALAKIPQKVLWRFDGNKPDALGLNTRLYKWIPQN
```

```
                   ***********::** : *******:********************:**************

2912330CD1         DLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTM
g4507823           DLLGHPKTRAFITHGGANGIYEAIYHGIPMVGIPLFFDQPDNIAHMKAKGAAVRLDFNTM
                   ********:******* **********:****:*:* ** *************.::*:**

2912330CD1         TSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAA
g4507823           SSTDLLNALKTVINDPLYKENIMKLSRIQHDQPVKPLDRAVFWIEFVMPHKGAKHLRVAA
                   :* ***.**:***.*. **** *:****:******************** ******** **

2912330CD1         HDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE
g4507823           HDLTWFQYHSLDVIGFLLACVATVIFIITKFCLFCFWKFARKGKKGKRD
                   *******::*:*******:****.**::**  **.  ** :. *  **:
```

Submit sequences to: BLAST2 Submit Reset

IncyteGenomics

Exhibit B
with Response dated 4/26/04
In USSN: 09/980,729

SeqServer®
biology in silico

BLAST2 Search Results

Sequences | Help

Retrieval | BLAST2 | FASTA | ClustalW | GCG Assembly | Phrap | Translation
BLAST2 Manual

 SeqServer Version 4.6 Jan 2002

Program: blastp
Sequence ID(s):
☐ 2912330CD1 vs. gb137mamp, gb137rodp

NCBI-BLASTP 2.0.10 [Aug-26-1999]

Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query= 2912330CD1
(529 letters)

Database: gb137mamp
27,779 sequences; 6,697,212 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
☑ g165799 UGT2B14; UDP-glucuronosyltransferase	779	0.0
☑ g2444022 UGT2B13; UDP-glucuronosyltransferase; EC 2.4.1.17	758	0.0
☑ g165797 UGT2B13; UDP-glucuronosyltransferase	758	0.0
☑ g165801 UGT2C1; UDP-glucuronosyltransferase	623	e-179
☑ g2773068 UGT1A; UDP-glucuronosyltransferase; EC 2.4.1.17	465	e-131
☑ g2842546 UGT1A1; UDP-glucuronosyltransferase	437	e-123
☑ g2773066 UGT1A; UDP-glucuronosyltransferase; EC 2.4.1.17	435	e-122
☑ g483789 UGT1-4; UDP-glucuronosyltransferase	428	e-120
☑ g2605508 UDP-glucuronosyltransferase	428	e-120
☑ g4760841 sheUGT1A6; UDP-glucuronosyltransferase	421	e-118

>g165799 UGT2B14; UDP-glucuronosyltransferase
Length = 530

Score = 779 bits (1990), Expect = 0.0
Identities = 360/530 (67%), Positives = 439/530 (81%), Gaps = 1/530 (0%)

```
Query: 1   MSMKWTSALLLI-QLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASS 59
           MS+K  S LLL+ QLSC F +GSCGKVLVWP +FS WMN+  ILDELV+RGHEV VL +S
```

```
Sbjct: 1   MSVKHVSVLLLLLQLSCCFRTGSCGKVLVWPMDFSLWMNLNVILDELVRRGHEVIVLRNS 60

Query: 60  ASISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPKDTFWSYFSQVQEIMWTF 119
           ASI  DP+  + +KFE +P++ TK + ED+    V  W    +++ W YFS +Q++   +
Sbjct: 61  ASIFIDPSKQANIKFETFPIAATKDDLEDLFVHYVSTWTNARQNSQWKYFSLLQKLFSEY 120

Query: 120 NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA 179
           +D     CK++V NK LM KLQESRFD++L+DA+ P GELLAELLKIPFVYSLRF+PGY
Sbjct: 121 SDSCENACKEVVFNKTLMTKLQESRFDILLSDAIGPCGELLAELLKIPFVYSLRFTPGYT 180

Query: 180 IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV 239
           +EK+SGGL  PPSYVP+++S+LS +MTF+ERV NM+ +LYF+FWFQ+F+ K+WDQFYSEV
Sbjct: 181 MEKYSGGLSVPPSYVPIILSDLSGKMTFMERVNNMLCMLYFDFWFQMFNKKRWDQFYSEV 240

Query: 240 LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS 299
           LGRP T SE + KAD+WLIR+YWD +FP P LPN++FVGGLHCKPAKPLPKEMEEFVQSS
Sbjct: 241 LGRPVTFSELVGKADMWLIRSYWDLEFPRPTLPNIQFVGGLHCKPAKPLPKEMEEFVQSS 300

Query: 300 GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ 359
           GE GVVVFSLGSMVSN +EERAN+IASA A++PQKV+WRFDG KP+TLG NTR+Y WIPQ
Sbjct: 301 GEEGVVVFSLGSMVSNMTEERANLIASAFAQLPQKVIWRFDGQKPETLGPNTRIYDWIPQ 360

Query: 360 NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT 419
           NDLLGHPKTKAF+THGG NGIYEAI+HG+PMVG+P+FG+Q DNIAHM AKGAA+ +N+KT
Sbjct: 361 NDLLGHPKTKAFVTHGGANGIYEAIHHGIPMVGLPLFGEQPDNIAHMTAKGAAIRLNWKT 420

Query: 420 MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA 479
           M+SEDLL AL+TVI D SYKEN M LS IHHDQP+KPLDRAVFWIE+VMRHKGAKHLR A
Sbjct: 421 MSSEDLLNALKTVINDPSYKENVMTLSSIHHDQPMKPLDRAVFWIEYVMRHKGAKHLRVA 480

Query: 480 AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 529
           AHDLTWFQ++S+DV+GFL++C A  IFL  K +LF  QK  K  K +KR+
Sbjct: 481 AHDLTWFQYHSLDVVGFLVSCAAFLIFLVIKSYLFVYQKLVKIGKKQKRD 530


>g2444022 UGT2B13; UDP-glucuronosyltransferase; EC 2.4.1.17
          Length = 523

 Score =  758 bits (1937), Expect = 0.0
 Identities = 358/522 (68%), Positives = 422/522 (80%), Gaps = 1/522 (0%)

Query: 9   LLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSASISFDPNS 68
           LLL+QLSC FSSGSCGKVLVWP EFSHWMN+KTILD LVQRGH VTVL SSASI  + N
Sbjct: 2   LLLLQLSCCFSSGSCGKVLVWPMEFSHWMNMKTILDALVQRGHAVTVLRSSASILVNSND 61

Query: 69  PSTLKFEVYPVSLTKTEFEDIIKQ-LVKRWAELPKDTFWSYFSQVQEIMWTFNDILRKFC 127
            S + FE +P + TK E E      L K   ++ KD  W YF   Q+    ++D     C
Sbjct: 62  ESGITFETFPTTSTKDEMEAFFMYWLNKLTNDVSKDALWEYFQTWQKFFMEYSDNYENIC 121

Query: 128 KDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYAIEKHSGGL 187
           KD+V NKK+M KLQESRFDVVLAD + P GELLAELL  P VYS+RF+PGY  EK+SGGL
Sbjct: 122 KDLVLNKKIMAKLQESRFDVVLADPIAPCGELLAELLNRPLVYSVRFTPGYTYEKYSGGL 181

Query: 188 LFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEVLGRPTTLS 247
           LFPPSYVPV+MS+LS QMTF+ERVKNM+++LYF+FWFQ+ ++K+WDQF SEVLGRP T S
Sbjct: 182 LFPPSYVPVIMSDLSGQMTFMERVKNMLWMLYFDFWFQMLNVKRWDQFCSEVLGRPVTFS 241

Query: 248 ETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSSGENGVVVF 307
           E + KA+IWLIR+YWD +FP PLLPN  FVGGLHCKPA+PLPKEMEEFVQSSGE GVVVF
Sbjct: 242 ELVGKAEIWLIRSYWDLEFPRPLLPNSYFVGGLHCKPAQPLPKEMEEFVQSSGEEGVVVF 301

Query: 308 SLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPK 367
           SLGSM+SN +EERANVIAS LA++PQKVLW+FDG KPD LG NT+LYKWIPQNDLLGH
Sbjct: 302 SLGSMISNLTEERANVIASTLAQLPQKVLWKFDGKKPDNLGTNTQLYKWIPQNDLLGHTV 361

Query: 368 TKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLR 427
           +KAFITHGG NG++EAIYHG+PMVG+P+F DQ DN+AHM+AKGAA+ +++KTM+S D L
```

```
Sbjct: 362 SKAFITHGGANGVFEAIYHGIPMVGLPLFADQHDNLAHMRAKGAAIRLDWKTMSSSDFLN 421

Query: 428 ALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQ 487
           AL+TVI D SYKE AM LSRIHHDQP+KPLD+A+FWIEFVMRHKGAKHLR AAHDLTWFQ
Sbjct: 422 ALKTVINDPSYKEKAMTLSRIHHDQPMKPLDQAIFWIEFVMRHKGAKHLRVAAHDLTWFQ 481

Query: 488 HYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 529
           ++S+DVIGFLL C+    +L  KC+L   Q    T K +KR+
Sbjct: 482 YHSLDVIGFLLACLTITTYLVIKCWLLVYQNILMTGKKKKRD 523


>g165797 UGT2B13; UDP-glucuronosyltransferase
          Length = 531

 Score =  758 bits (1936), Expect = 0.0
 Identities = 362/522 (69%), Positives = 423/522 (80%), Gaps = 1/522 (0%)

Query: 9   LLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSASISFDPNS 68
           LLL+QLSC FSSGSCGKVLVWP EFSHWMN+KTILD LVQ+GHEVTVL SSASI    N+
Sbjct: 10  LLLLQLSCCFSSGSCGKVLVWPMEFSHWMNMKTILDALVQQGHEVTVLRSSASIVIGSNN 69

Query: 69  PSTLKFEVYPVSLTKTEFEDIIKQ-LVKRWAELPKDTFWSYFSQVQEIMWTFNDILRKFC 127
            S +KFE +  S  K E E+       K    +  +++W  FS  + ++  ++DI    C
Sbjct: 70  ESGIKFETFHTSYRKDEIENFFMDWFYKMIYNVSIESYWETFSLTKMVILKYSDICEDIC 129

Query: 128 KDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYAIEKHSGGL 187
           K+++ NKKLM KLQESRFDVVLAD V P GELLAELLKIP VYSLR   GY ++KH GGL
Sbjct: 130 KEVILNKKLMTKLQESRFDVVLADPVSPGGELLAELLKIPLVYSLRGFVGYMLQKHGGGL 189

Query: 188 LFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEVLGRPTTLS 247
           L PPSYVPV+MS L  QMTF+ERV+N++ VLYF+FWF  F+ K+WDQFYSEVLGRP T
Sbjct: 190 LLPPSYVPVMMSGLGSQMTFMERVQNLLCVLYFDFWFPKFNEKRWDQFYSEVLGRPVTFL 249

Query: 248 ETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSSGENGVVVF 307
           E M KAD+WLIR+YWD +FP PLLPN +F+GGLHCKPAKPLP+EME+FVQSSGE GVVVF
Sbjct: 250 ELMGKADMWLIRSYWDLEFPRPLLPNFDFIGGLHCKPAKPLPQEMEDFVQSSGEEGVVVF 309

Query: 308 SLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPK 367
           SLGSM+SN +EERANVIASALA++PQKVLWRF+G KPD LG NTRLYKWIPQNDLLGHPK
Sbjct: 310 SLGSMISNLTEERANVIASALAQLPQKVLWRFEGKKPDMLGSNTRLYKWIPQNDLLGHPK 369

Query: 368 TKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLR 427
           TKAFITHGG NG++EAIYHG+PMVG+P+FGDQLDNI +MKAKGAAV++N KTM+S DLL
Sbjct: 370 TKAFITHGGANGVFEAIYHGIPMVGLPLFGDQLDNIVYMKAKGAAVKLNLKTMSSADLLN 429

Query: 428 ALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQ 487
           AL+TVI D SYKENAM LSRIHHDQP+KPLDRAVFWIE+VMRHKGAKHLR AAHDLTW+Q
Sbjct: 430 ALKTVINDPSYKENAMTLSRIHHDQPMKPLDRAVFWIEYVMRHKGAKHLRVAAHDLTWYQ 489

Query: 488 HYSIDVIGFLLTCVAITTYLIVKCCLLVYRYVLGAGKKKKRD 531
           ++S+DVIGFLL CVA   +L   +  L   +    K +KR+
Sbjct: 490 YHSLDVIGFLLACVAITTYLIVKCCLLVYRYVLGAGKKKKRD 531


>g165801 UGT2C1; UDP-glucuronosyltransferase
          Length = 502

 Score =  623 bits (1590), Expect = e-179
 Identities = 301/499 (60%), Positives = 383/499 (76%), Gaps = 2/499 (0%)

Query: 32  EFSHWMNIKTILDELVQRGHEVTVLASSASISFDPNSPSTLKFEVYPVSLTK-TEFEDII 90
           +FSHW+N+K IL+EL  RGHE+TVL  S S+  D ++       EV  + +TK T  E++
Sbjct: 5   DFSHWINLKVILEELQLRGHEITVLVPSPSLLLD-HTKIPFNVEVLQLQVTKETLMEELN 63

Query: 91  KQLVKRWAELPKDTFWSYFSQVQEIMWTFNDILRKFCKDIVSNKKLMKKLQESRFDVVLA 150
             L     ELP  ++W    ++ E+   F+  LR+ C   ++NK+L+ +L+ ++FD+ LA
Sbjct: 64  TVLYMSSFELPTLSWWKVLGKMVEMGKQFSKNLRRVCDSAITNKELLDRLKAAKFDICLA 123
```

```
Query: 151 DAVFPFGELLAELLKIPFVYSLRFSPGYAIEKHSGGLLFPPSYVPVVMSELSDQMTFIER 210
           D +   GEL+AELL IPFVYS RFS G  IE+   GL  P SYVP   S L+D M+F++R
Sbjct: 124 DPLAFCGELVAELLNIPFVYSFRFSIGNIIERSCAGLPTPSSYVPGSTSGLTDNMSFVQR 183

Query: 211 VKNMIYVLYFEFWFQIFDMKKWDQFYSEVLGRPTTLSETMAKADIWLIRNYWDFQFPHPL 270
           +KN +  L  +  F  F + +WD++YS+VLGR TT+ E M KA++WLIR+YWDF+FP P
Sbjct: 184 LKNWLLYLMNDMMFSHFMLSEWDEYYSKVLGRRTTICEIMGKAEMWLIRSYWDFEFPRPF 243

Query: 271 LPNVEFVGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVSNTSEERANVIASALAK 330
           LPN E+VGGLHCKPAKPLP+E+EEFVQSSG +GVVVF+LGSM+ N +EER+N+IASALA+
Sbjct: 244 LPNFEYVGGLHCKPAKPLPEELEEFVQSSGNDGVVVFTLGSMIQNLTEERSNLIASALAQ 303

Query: 331 IPQKVLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPM 390
           IPQKVLWR+ G KP TLG NTRL++WIPQNDLLGHPKT+AFITHGG NG+YEAIYHGVPM
Sbjct: 304 IPQKVLWRYTGKKPATLGPNTRLFEWIPQNDLLGHPKTRAFITHGGTNGLYEAIYHGVPM 363

Query: 391 VGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALRTVITDSSYKENAMRLSRIHH 450
           VG+P+FGDQ DNIA +KAKGAAV+++ + MT+  LL+AL+ VI + SYKENAM+LSRIHH
Sbjct: 364 VGIPLFGDQPDNIARVKAKGAAVDVDLRIMTTSSLLKALKDVINNPSYKENAMKLSRIHH 423

Query: 451 DQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQHYSIDVIGFLLTCVATAIFLFTK 510
           DQP+KPLDRAVFWIEFVMRHKGA+HLR AAHDLTWFQ+YS+DV+ FLLTCVAT IFL  K
Sbjct: 424 DQPLKPLDRAVFWIEFVMRHKGARHLRVAAHDLTWFQYYSLDVVVFLLTCVATIIFLAKK 483

Query: 511 CFLFSCQKFNKTRKIEKRE 529
           C LF  ++F KT    KRE
Sbjct: 484 CCLFFYRRFCKTGNKRKRE 502


>g2773068 UGT1A; UDP-glucuronosyltransferase; EC 2.4.1.17
          Length = 533

 Score =  465 bits (1183), Expect = e-131
 Identities = 238/518 (45%), Positives = 330/518 (62%), Gaps = 7/518 (1%)

Query: 11  LIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSASISFDPNSPS 70
           L+   C       GKVLV PT+ SHW+++K  L EL  RGHE+ V++    ++
Sbjct: 15  LLLFLCVGPRAEGGKVLVLPTDGSHWLSMKKALQELHARGHEIVVVSPEVNLHIKKEDFF 74

Query: 71  TLKFEVYPVSLTKTEFEDIIKQLVKRWAELPKDTFWSYFSQVQEIMWTFNDILRKFCKDI 130
           TL+   Y +S T+ EF D    L   +    K  F   F ++ E + T + I ++ CK++
Sbjct: 75  TLR--TYAISYTQEEFNDFF--LGHSYLVFEKGHFLKMFLKIMENLKTASFIFQRSCKEL 130

Query: 131 VSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYAIEKHSGGLLFP 190
           + NK+L+  L  S FDVVL D V+P G +LA+ L +P V+ LR  P   ++         P
Sbjct: 131 MHNKELIGHLNSSSFDVVLTDPVYPCGAVLAKYLSLPAVFFLRSVP-CDLDFEGTQCPNP 189

Query: 191 PSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEVLGRPTTLSETM 250
            SY+P +++  SD MTF++RVKNM+Y L  ++   I     +    SE+L R  ++ +
Sbjct: 190 SSYIPRLLTMNSDHMTFLQRVKNMLYPLSLKYICHIA-FTPYASLASELLQREVSVVDVF 248

Query: 251 AKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLG 310
           + A +WL R  +   +P P++PN+ F+GG++C   KPL +E E +V +SGE+G+VVFSLG
Sbjct: 249 SSASMWLFRGDFVLDYPRPVMPNMVFIGGINCANRKPLSQEFEAYVNASGEHGIVVFSLG 308

Query: 311 SMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTKA 370
           SMVS   +E+A  IA AL KIPQ VLWR+ G  P  L  NT L KW+PQNDLLGHPK +A
Sbjct: 309 SMVSAIPKEKAMEIADALGKIPQTVLWRYTGTPPPNLAKNTILVKWLPQNDLLGHPKARA 368

Query: 371 FITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALR 430
           FITH G +GIYE I +GVPMV +P+FGDQ+DN   M+ +GA + +N   MTSEDL   L+
Sbjct: 369 FITHSGSHGIYEGICNGVPMVMLPLFGDQMDNAKRMETRGAGLTLNVLEMTSEDLANGLK 428

Query: 431 TVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQHYS 490
           VI D SYKEN MRLS +H D+P++PLD AVFW+EFVMRHKGA HLR AAHDLTW+Q++S
Sbjct: 429 AVINDKSYKENIMRLSSLHKDRPIEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQYHS 488
```

```
Query: 491 IDVIGFLLTCVATAIFLFTKCFLFSCQK-FNKTRKIEK 527
          +DVIGFLL  V   +F+  KC  F C+K F +  +++K
Sbjct: 489 VDVIGFLLAIVLGIVFITYKCCAFGCRKCFGRKGRVKK 526


>g2842546 UGT1A1; UDP-glucuronosyltransferase
          Length = 533

 Score =  437 bits (1111), Expect = e-123
 Identities = 237/521 (45%), Positives = 322/521 (61%), Gaps = 9/521 (1%)

Query: 8   ALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSASISFDPN 67
          +LLL  L+   S G  GK+L+ P + SHW+++  ++  L QRGH+V V+A  AS+
Sbjct: 14  SLLLCALNPLLSQG--GKLLLVPMDGSHWLSLFGVIQRLHQRGHDVVVVAPEASVYIKEG 71

Query: 68  SPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPKDTFWSYFSQVQEIMWTFNDILRKFC 127
          +  TLK   YPV   + + E     L     E  K  F        + +   + +L   C
Sbjct: 72  AFYTLKS--YPVPFRRVDVEASFTGLGLGIFE--KKPFLRRVVATYKRVKKDSALLLSAC 127

Query: 128 KDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYAIEKHSGGL 187
            ++ N++LM  L ES FD +L D   P G ++A  L  P V+ L   P   ++
Sbjct: 128 SHLLYNEELMASLAESGFDAMLTDPFLPCGPIVALRLAWPVVFFLNSLP-CGLDFQGTRC 186

Query: 188 LFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEVLGRPTTLS 247
            PPSYVP V+S  SD MTF++RVKNM+ +L  E +         +    SEVL +  T+
Sbjct: 187 PSPPSYVPRVLSLNSDHMTFLQRVKNML-ILGSEGFLCNVVYSPYASLASEVLQKDVTVQ 245

Query: 248 ETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSSGENGVVVF 307
          + M  A +WL R+ +   +  P++PN+ F+GG++C    PL +E E +V +SGE+G+VVF
Sbjct: 246 DLMGSASVWLFRSDFVKDYSRPIMPNMVFIGGINCAGKNPLSQEFEAYVNASGEHGIVVF 305

Query: 308 SLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPK 367
          SLGSMVS   +E+A  IA AL KIPQ VLWR+ G  P  L  NT L KW+PQNDLLGHPK
Sbjct: 306 SLGSMVSEIPKEKAMEIADALGKIPQTVLWRYTGTPPPNLAKNTILVKWLPQNDLLGHPK 365

Query: 368 TKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLR 427
           +AFITH G +GIYE I +GVPMV +P+FGDQ+DN   M+ +GA + +N   MTSEDL
Sbjct: 366 ARAFITHSGSHGIYEGICNGVPMVMLPLFGDQMDNAKRMETRGAGLTLNVLEMTSEDLAN 425

Query: 428 ALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQ 487
          AL+ VI D SYKEN MRLS +H D+P++PLD AVFW+EFVMRHKGA HLR AAHDLTW+Q
Sbjct: 426 ALKAVINDKSYKENIMRLSSLHKDRPIEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQ 485

Query: 488 HYSIDVIGFLLTCVATAIFLFTKCFLFSCQK-FNKTRKIEK 527
          ++S+DVIGFLL  V   +F+  KC  F C+K F K  +++K
Sbjct: 486 YHSVDVIGFLLAIVLGIVFITYKCCAFGCRKCFGKKGRVKK 526


>g2773066 UGT1A; UDP-glucuronosyltransferase; EC 2.4.1.17
          Length = 533

 Score =  435 bits (1107), Expect = e-122
 Identities = 235/521 (45%), Positives = 322/521 (61%), Gaps = 9/521 (1%)

Query: 8   ALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSASISFDPN 67
          +LLL  L+   S G  GK+L+ P + SHW+++  ++  L QRGH+V V+A  AS+
Sbjct: 14  SLLLCALNPLLSQG--GKLLLVPMDGSHWLSLFGVIQRLHQRGHDVVVVAPEASVYIKEG 71

Query: 68  SPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPKDTFWSYFSQVQEIMWTFNDILRKFC 127
          +  TLK   YPV   + + E     L     E  K  F        + +   + +L   C
Sbjct: 72  AFYTLKS--YPVPFRREDVEASFTGLGLGVFE--KKPFLQRVVATYKRVKKDSALLLSAC 127

Query: 128 KDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYAIEKHSGGL 187
            ++ N++LM  L ES FD +L D   P G ++A  L +P V+ L   P   ++
Sbjct: 128 SHLLYNEELMASLAESGFDAMLTDPFLPCGPIVALRLALPVVFFLNSLP-CGLDFQGTRC 186
```

```
Query: 188 LFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEVLGRPTTLS 247
            PPSYVP V+S  SD MTF++RVKNM+ +L  E +         +    SEVL +  T+
Sbjct: 187 PSPPSYVPRVLSLNSDHMTFLQRVKNML-ILGSEGFLCNVVYSPYASLASEVLQKDVTVQ 245

Query: 248 ETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSSGENGVVVF 307
           + M  A +WL R+ +   +  P++PN+ F+GG++C    PL +E E +V +SGE+G+VVF
Sbjct: 246 DLMGSASVWLFRSDFVKDYSRPIMPNMVFIGGINCAGKNPLSQEFEAYVNASGEHGIVVF 305

Query: 308 SLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPK 367
           SLGSMVS   +E+A  IA AL KIPQ VLWR+ G  P  L  NT L KW+PQNDLLGHPK
Sbjct: 306 SLGSMVSAIPKEKAMEIADALGKIPQTVLWRYTGTPPPNLAKNTILVKWLPQNDLLGHPK 365

Query: 368 TKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLR 427
            +AFITH G +GIYE I +GVPMV +P+FGDQ+DN   M+ +GA + +N   MTSEDL
Sbjct: 366 ARAFITHSGSHGIYEGICNGVPMVMLPLFGDQMDNAKRMETRGAGLTLNVLEMTSEDLAN 425

Query: 428 ALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQ 487
            L+ VI D SYKEN MRLS +H D+P++PLD AVFW+EFVMRHKGA HLR AAHDLTW+Q
Sbjct: 426 GLKAVINDKSYKENIMRLSSLHKDRPIEPLDLAVFWVEFVMRHKGAPHLRPAAHDLTWYQ 485

Query: 488 HYSIDVIGFLLTCVATAIFLFTKCFLFSCQK-FNKTRKIEK 527
           ++S+DVIGFLL  V   +F+  KC  F C+K F +  +++K
Sbjct: 486 YHSVDVIGFLLAIVLGIVFITYKCCAFGCRKCFGRKGRVKK 526
```

>g483789 UGT1-4; UDP-glucuronosyltransferase
Length = 532

Score = 428 bits (1089), Expect = e-120
Identities = 227/518 (43%), Positives = 329/518 (62%), Gaps = 8/518 (1%)

```
Query: 11  LIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSASISFDPNSPS 70
           L+ L C       GKVLV P + S W++++ ++ ++  RGH+V VL    ++
Sbjct: 15  LLLLLCVLPWAEGGKVLVVPMDGSPWLSLREVVRDVHARGHQVLVLGPEVTMHIKGEDFF 74

Query: 71  TLKFEVYPVSLTKTEFEDIIKQLVKRWAELPKDTFWSYFSQVQEIMWTFNDILRKFCKDI 130
           TL  + Y    +K EF+ ++++  +   + P+ +  +    ++ +   F+ +  + C ++
Sbjct: 75  TL--QTYATPYSKEEFDQLMQRNYQMIFK-PQHSLKTLLETMENLK-KFSMLCSRSCWEL 130

Query: 131 VSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYAIEKHSGGLLFP 190
           + NK L+K L ES FDVVL D +   G LLA+ L +P V+ LRF     ++        P
Sbjct: 131 LHNKPLIKHLNESSFDVVLTDPLDLCGALLAKYLSVPSVFLLRFIL-CDLDFEGTQCPNP 189

Query: 191 PSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEVLGRPTTLSETM 250
            SY+P +++  SD M+F++RVKNM+Y L  ++   I     +    SE+  R  +L + +
Sbjct: 190 SSYIPRMLTMNSDHMSFLQRVKNMLYPLMMKYTCHI-SYDPYASLASELFQREVSLVDIL 248

Query: 251 AKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLG 310
           + A +WL R  +   +P P++PN+ F+GG++C    KPL +E E +V +SGE+G+VVFSLG
Sbjct: 249 SHASVWLFREDFVLDYPRPIMPNMVFIGGINCANRKPLSQEFEAYVNASGEHGIVVFSLG 308

Query: 311 SMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTKA 370
           SMVS   E++A  IA AL KIPQ VLWR+ G++P  L  NT L KW+PQN LLGHPKT+A
Sbjct: 309 SMVSEIPEKKAMEIADALGKIPQTVLWRYTGSRPSNLAKNTYLVKWLPQNVLLGHPKTRA 368

Query: 371 FITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALR 430
           FITH G +GIYE I +GVPMV +P+FGDQ+DN   ++ +GA V +N   MTS+DL  AL+
Sbjct: 369 FITHSGSHGIYEGICNGVPMVMLPLFGDQMDNAKRIETRGAGVTLNVLEMTSDDLANALK 428

Query: 431 TVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQHYS 490
           TVI D SYKEN MRLS +H D+PV+PLD AVFW+EFVMRHKGA   R AAHDLTW+Q++S
Sbjct: 429 TVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAAP-RPAAHDLTWYQYHS 487

Query: 491 IDVIGFLLTCVATAIFLFTKCFLFSCQK-FNKTRKIEK 527
           +DVIGFLL  V T  F+  KC  F+  K F K  +++K
Sbjct: 488 LDVIGFLLAIVLTVAFVTFKCCAFAWGKCFGKKGRVKK 525
```

```
>g2605508 UDP-glucuronosyltransferase
          Length = 529

 Score =  428 bits (1089), Expect = e-120
 Identities = 228/509 (44%), Positives = 314/509 (60%), Gaps = 17/509 (3%)

Query: 25  KVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSASISFDPNSPSTLKFEVYPVSLTKT 84
           ++LV P + SHW+++K I++ L ++GHE+ V+    ++    +   T K  ++PV   +
Sbjct: 25  RLLVVPQDGSHWLSMKDIVEHLSEKGHEIVVVVPEVNLLLQESKHYTRK--IHPVPFNQE 82

Query: 85  EFEDIIKQLVKRWAELPKDTF---WSYFSQVQEIMWTFNDILRKF--CKDIVSNKKLMKK 139
           E E        R+    K  F   W   + V E       I   F  C+ ++ +   ++
Sbjct: 83  ELE-------ARYRSFGKHHFSPRWLVTAPVVEYRNNMIVINMYFLNCQSLLRHSDTLRF 135

Query: 140 LQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYAIEKHSGGLLFPPSYVPVVMS 199
           L+E++FD +  D   P G +LAE L +P VY  R  P  A+E         P SYVP   +
Sbjct: 136 LRENKFDALFTDPALPCGVILAEYLNLPSVYLFRGFP-CALENTFTRTPSPLSYVPRYYT 194

Query: 200 ELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEVLGRPTTLSETMAKADIWLIR 259
           + SD MTF++RV N + V Y E          K++    EVLGR   L     KA IWL+R
Sbjct: 195 QFSDHMTFLQRVGNFL-VNYLENILLYALYSKYEDLAGEVLGRQVHLPALYRKASIWLLR 253

Query: 260 NYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVSNTSEE 319
             + F++P P++PN   +GG  CK    L +E E +V +SGE+G+VVFSLGSMVS   E+
Sbjct: 254 YDFVFEYPRPVMPNTVLIGGSSCKKQGVLSQEFEAYVNASGEHGIVVFSLGSMVSEIPEQ 313

Query: 320 RANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTKAFITHGGMNG 379
           +A  IA AL KIPQ VLWR+ G  P  L  NT+L KW+PQNDLLGHPKT+AFITH G +G
Sbjct: 314 KAMEIADALGKIPQTVLWRYTGTPPPNLAKNTKLVKWLPQNDLLGHPKTRAFITHSGSHG 373

Query: 380 IYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALRTVITDSSYK 439
           IYE I +GVPMV +P+FGDQ+DN   M+ +GA V +N   M+SEDL +AL+ VI + +YK
Sbjct: 374 IYEGICNGVPMVMMPLFGDQMDNAKRMETRGAGVTLNVLEMSSEDLEKALKAVINEKTYK 433

Query: 440 ENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQHYSIDVIGFLLT 499
           EN MRLSR+H D+P++PLD AVFW+EFVMRHKGA HLR AAHDLTW+Q++S+DVIGFLL
Sbjct: 434 ENIMRLSRLHKDRPIEPLDLAVFWVEFVMRHKGASHLRPAAHDLTWYQYHSLDVIGFLLA 493

Query: 500 CVATAIFLFTKCFLFSCQK-FNKTRKIEK 527
              T IF+  K   F+ +K F K  +++K
Sbjct: 494 VTLTVIFITFKACAFAFRKCFGKKERVKK 522


>g4760841 sheUGT1A6; UDP-glucuronosyltransferase
          Length = 531

 Score =  421 bits (1072), Expect = e-118
 Identities = 221/504 (43%), Positives = 312/504 (61%), Gaps = 7/504 (1%)

Query: 25  KVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSASISFDPNSPSTLKFEVYPVSLTKT 84
           ++LV P + SHW+++K I + L ++GHE+ V+    ++    +   T +  ++PV   +
Sbjct: 27  RLLVVPQDGSHWLSMKDITERLSEKGHEIVVVVPKVNLLLQESKHYTRR--IHPVPYDQE 84

Query: 85  EFEDIIKQLVKRWAELPKDTFWSYFSQVQEIMWTFNDILRKFCKDIVSNKKLMKKLQESR 144
           E E   +   K      P+    +  + + M   N       C+ ++ +   ++KLQ S+
Sbjct: 85  ELEARYRSFGKHHFS-PRWLVTAPMVEYRNNMIVINMYFLN-CQSLLRHSGTLRFLRESK 142

Query: 145 FDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYAIEKHSGGLLFPPSYVPVVMSELSDQ 204
           FD +  D   P G +LAE L +P VY  R  P  A+E         P SYVP   ++ SD+
Sbjct: 143 FDALFTDPALPCGVILAEYLNLPSVYLFRGFP-CALENTFTRTPSPLSYVPRYYTQFSDK 201

Query: 205 MTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEVLGRPTTLSETMAKADIWLIRNYWDF 264
           MTF++RV N + V Y E          K++    EVLGR   L     KA IWL+R  + F
Sbjct: 202 MTFLQRVANFL-VSYLENILLYALYSKYEDLAEEVLGRQVHLPALYQKASIWLLRYDFVF 260

Query: 265 QFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSSGENGVVVFSLGSMVSNTSEERANVI 324
```

```
           ++P P++PN+ F+GG   K    LP+E E +V +SGE+G+V+FSLGSMVS   E++A  I
Sbjct: 261 EYPRPVMPNMVFIGGSAAKKQGILPREFEAYVNASGEHGIVIFSLGSMVSEIPEQKAMEI 320

Query: 325 ASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQNDLLGHPKTKAFITHGGMNGIYEAI 384
           A AL KIPQ VLWR+ G  P  L  NT+L KW+PQNDLLG PKT+AFITH G +G+YE I
Sbjct: 321 ADALGKIPQTVLWRYTGTPPPNLAKNTKLVKWLPQNDLLGQPKTRAFITHSGSHGVYEGI 380

Query: 385 YHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMTSEDLLRALRTVITDSSYKENAMR 444
            +GVPMV +P+FGDQ+DN   M+ +GA + +N   M+S DL  AL+ VI + SYKEN MR
Sbjct: 381 CNGVPMVMMPLFGDQMDNAERMETRGAGITLNVLEMSSGDLENALKAVINEKSYKENIMR 440

Query: 445 LSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAHDLTWFQHYSIDVIGFLLTCVATA 504
           LSR+H D+P++PLD AVFW+EFVMRHKGA HLR AAHDLTW+Q++S+DVIGFLL    T
Sbjct: 441 LSRLHKDRPIEPLDLAVFWVEFVMRHKGASHLRPAAHDLTWYQYHSLDVIGFLLAVTLTV 500

Query: 505 IFLFTKCFLFSCQK-FNKTRKIEK 527
           IF+  K   F+ +K F K  +++K
Sbjct: 501 IFITFKACAFTFRKCFGKKERVKK 524


  Database: gb137mamp
    Posted date:  Sep 11, 2003 11:25 AM
  Number of letters in database: 6,697,212
  Number of sequences in database:  27,779

Lambda     K      H
   0.324    0.138    0.428

Gapped
Lambda     K      H
   0.270   0.0470    0.230


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 7447897
Number of Sequences: 27779
Number of extensions: 323061
Number of successful extensions: 685
Number of sequences better than 10.0: 33
Number of HSP's better than 10.0 without gapping: 18
Number of HSP's successfully gapped in prelim test: 15
Number of HSP's that attempted gapping in prelim test: 623
Number of HSP's gapped (non-prelim): 35
length of query: 529
length of database: 6,697,212
effective HSP length: 46
effective length of query: 483
effective length of database: 5,419,378
effective search space: 2617559574
effective search space used: 2617559574
T: 11
A: 40
X1: 15 ( 7.0 bits)
X2: 38 (14.8 bits)
X3: 64 (24.9 bits)
S1: 41 (22.0 bits)
```

```
NCBI-BLASTP 2.0.10 [Aug-26-1999]


Reference: Altschul, Stephen F., Thomas L. Madden, Alejandro A. Schaffer,
Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997),
```

"Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402.

Query= 2912330CD1
(529 letters)

Database: gb137rodp
74,095 sequences; 25,169,402 total letters

Searching.....done

Sequences producing significant alignments:	Score (bits)	E Value
g20071113 RIKEN cDNA 1300012D20	782	0.0
g20380046 expressed sequence AI788959	779	0.0
g207581 UDP-glucuronosyltransferase (EC 2.4.1.17)	772	0.0
g207569 UDP glucuronosyltransferase-2	772	0.0
g18146841 UGT2B21; UDP-glucuronosyltransferase 2B21; EC 2.7.1	771	0.0
g458395 UDP-glucuronosyltransferase; EC 2.4.1.17	755	0.0
g458397 UDP-glucuronosyltransferase; EC 2.4.1.17	747	0.0
g20381430 UDP-glucuronosyltransferase 2 family, member 5	747	0.0
g55120 unnamed protein product; UDP-glucuronosyltransferase p	745	0.0
g15929692 RIKEN cDNA 9430041C03	742	0.0

>g20071113 RIKEN cDNA 1300012D20
Length = 529

Score = 782 bits (1998), Expect = 0.0
Identities = 369/528 (69%), Positives = 431/528 (80%), Gaps = 1/528 (0%)

```
Query: 1   MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA 60
           MSMK  S  LLIQ  CY   G+CGKVLVWPTE+SHW+N+K ILDELVQRGH+VTVL SSA
Sbjct: 1   MSMKQASVFLLIQFICYIRPGACGKVLVWPTEYSHWINMKIILDELVQRGHDVTVLISSA 60

Query: 61  SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMWTF 119
           SI   P++ S++ FE+Y   L+K + E   ++ V  W  EL K  FW+ +S++Q+I   +
Sbjct: 61  SILIGPSNESSINFEIYSAPLSKDDLEYAFEKWVGNWTYELKKLPFWTSYSKLQKISSEY 120

Query: 120 NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA 179
           +D++  FCK +V NK LMKKLQ S+FDVVLADA+ P GELL+ELLK P VYSLRF PGY
Sbjct: 121 SDMIESFCKAVVWNKSLMKKLQGSKFDVVLADALVPCGELLSELLKTPLVYSLRFCPGYK 180

Query: 180 IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV 239
            EK+SGGL  PPSYVPVV+SELSD MTF ERVKNM+ VL F+FWFQ F+ K W+QFYS+V
Sbjct: 181 CEKYSGGLPLPPSYVPVVLSELSDHMTFAERVKNMLQVLLFDFWFQTFNEKSWNQFYSDV 240

Query: 240 LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS 299
           LGRPTTL+E M KADIWL+R +WD +FPHP LPN +FVGGLHCKPAKPLPKEMEEFVQSS
Sbjct: 241 LGRPTTLTEMMGKADIWLVRTFWDLKFPHPFLPNFDFVGGLHCKPAKPLPKEMEEFVQSS 300

Query: 300 GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ 359
           GE+GVVVFSLGSMV N  EE+ANV+ASALA+IPQKVLWRFDG KPDTLG NTRLYKWIPQ
Sbjct: 301 GEHGVVVFSLGSMVKNIKEEKANVVASALAQIPQKVLWRFDGKKPDTLGSNTRLYKWIPQ 360

Query: 360 NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT 419
           NDLLGHPKTKAFI HGG NGIYEAIYHG+P+VG+P+FGDQ DNI H+ AKGAAV ++F T
Sbjct: 361 NDLLGHPKTKAFIAHGGTNGIYEAIYHGIPIVGIPLFGDQPDNINHIVAKGAAVRVDFDT 420

Query: 420 MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA 479
           M++ DLL AL+TVI D SYKENAMRLSRIHHDQP+KPLDRAVFWIE+VMR+KGAKHLR A
Sbjct: 421 MSTTDLLTALKTVINDPSYKENAMRLSRIHHDQPMKPLDRAVFWIEYVMRNKGAKHLRPA 480
```

```
Query: 480 AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEK 527
           HDLTWFQ++S+DVIGFLL CV   +F+  KC LF C K     K +K
Sbjct: 481 LHDLTWFQYHSLDVIGFLLVCVVAVVFIIAKCCLFCCHKTANMGKKKK 528
```

>g20380046 expressed sequence AI788959
Length = 532

```
 Score =  779 bits (1990), Expect = 0.0
 Identities = 365/532 (68%), Positives = 443/532 (82%), Gaps = 3/532 (0%)

Query: 1   MSMKWTSALLLI--QLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLAS 58
           M +K T+ALLL+  QLS +F SG+ GKVLVWP EFSHW+N+KTILDEL+++GHEV VL
Sbjct: 1   MPVKMTAALLLLLLQLSGFFGSGTGGKVLVWPMEFSHWLNLKTILDELLKKGHEVMVLRP 60

Query: 59  SASISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMW 117
           SAS+S++ ++ S ++FE YP S + +E E+I  + +K++  ELPK +FW YF  +QE++W
Sbjct: 61  SASLSYEVDNTSAIEFETYPTSYSLSELEEIFWESLKKYIYELPKQSFWGYFLMLQEMVW 120

Query: 118 TFNDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPG 177
             +      CKD+V NK+LM KLQ+SRFDV+LAD   P G+LLAE+LKIP VYSLRF PG
Sbjct: 121 VDSKYFESLCKDVVFNKELMTKLQKSRFDVILADPFIPCGDLLAEVLKIPLVYSLRFFPG 180

Query: 178 YAIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYS 237
              EK+SGGL  PPSYVPVVMSELSD+MTF+ERV+N+IY+L F+FWFQ F+ K W+Q Y+
Sbjct: 181 STYEKYSGGLPLPPSYVPVVMSELSDRMTFMERVRNVIYMLCFDFWFQTFNEKNWNQLYT 240

Query: 238 EVLGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQ 297
           EVLGRPTTLSETMAKADIWLIR YWD +FPHP+LPN +F+GGLHC+PAKPLPKE+E+FVQ
Sbjct: 241 EVLGRPTTLSETMAKADIWLIRTYWDLEFPHPVLPNFDFIGGLHCRPAKPLPKEIEDFVQ 300

Query: 298 SSGENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWI 357
           SSGE+GVVVFSLGSMV + +EERANVIA+ LA+IPQKVLWRF+G KP+TLG NTRLYKWI
Sbjct: 301 SSGEHGVVVFSLGSMVGSITEERANVIAAGLAQIPQKVLWRFEGKKPETLGSNTRLYKWI 360

Query: 358 PQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINF 417
           PQNDLLGH KT+AFITHGG NGIYEAIYHG+P+VG+P+FGDQ DNI H+KAKGAAV ++F
Sbjct: 361 PQNDLLGHSKTRAFITHGGTNGIYEAIYHGIPVVGIPLFGDQYDNIVHLKAKGAAVRLDF 420

Query: 418 KTMTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLR 477
            TM+S DL  AL+TV  D SYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLR
Sbjct: 421 LTMSSTDLHTALKTVTNDPSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLR 480

Query: 478 SAAHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 529
           AAHDL+W Q++S+DV+GFLL CV T +F+  KC LF CQK  K  + +K E
Sbjct: 481 VAAHDLSWVQYHSLDVLGFLLACVLTVMFILKKCCLFCCQKLTKAGRKKKGE 532
```

>g207581 UDP-glucuronosyltransferase (EC 2.4.1.17)
Length = 529

```
 Score =  772 bits (1972), Expect = 0.0
 Identities = 361/528 (68%), Positives = 429/528 (80%), Gaps = 1/528 (0%)

Query: 1   MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA 60
           MSMK TS  LLIQL CYF  G+CGKVLVWPTE+SHW+NIK IL+EL QRGHEVTVL SSA
Sbjct: 1   MSMKQTSVFLLIQLICYFRPGACGKVLVWPTEYSHWINIKIILNELAQRGHEVTVLVSSA 60

Query: 61  SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMWTF 119
           SI  +P   S++ FE+Y V L+K++ E    + +  W  +    + W+Y+S++Q++   +
Sbjct: 61  SILIEPTKESSINFEIYSVPLSKSDLEYSFAKWIDEWTRDFETLSIWTYYSKMQKVFNEY 120

Query: 120 NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA 179
           +D++   CK ++ NK LMKKLQ S+FDV+LADAV P GELLAELLK P VYSLRF PGY
Sbjct: 121 SDVVENLCKALIWNKSLMKKLQGSQFDVILADAVGPCGELLAELLKTPLVYSLRFCPGYR 180

Query: 180 IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV 239
```

```
            EK SGGL  PPSYVPVV+SELSD+MTF+ERVKNM+ +LYF+FWFQ F  K W QFYS+V
Sbjct: 181 CEKFSGGLPLPPSYVPVVLSELSDRMTFVERVKNMLQMLYFDFWFQPFKEKSWSQFYSDV 240

Query: 240 LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS 299
           LGRPTTL+E M KADIWLIR +WD +FPHP LPN +FVGGLHCKPAKPLP+EMEEFVQSS
Sbjct: 241 LGRPTTLTEMMGKADIWLIRTFWDLEFPHPFLPNFDFVGGLHCKPAKPLPREMEEFVQSS 300

Query: 300 GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ 359
           GE+GVVVFSLGSMV N +EE+ANV+ASALA+IPQKV+WRFDG KPDTLG NTRLYKWIPQ
Sbjct: 301 GEHGVVVFSLGSMVKNLTEEKANVVASALAQIPQKVVWRFDGKKPDTLGSNTRLYKWIPQ 360

Query: 360 NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT 419
           NDLLGHPKTKAF+ HGG NGIYEAIYHG+P+VG+P+F DQ DNI HM AKGAAV ++F
Sbjct: 361 NDLLGHPKTKAFVAHGGTNGIYEAIYHGIPIVGIPLFADQPDNINHMVAKGAAVRVDFSI 420

Query: 420 MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA 479
           +++  LL AL+ V+ D SYKENAMRLSRIHHDQPVKPLDRAVFWIE+VMRHKGAKHLRS
Sbjct: 421 LSTTGLLTALKIVMNDPSYKENAMRLSRIHHDQPVKPLDRAVFWIEYVMRHKGAKHLRST 480

Query: 480 AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEK 527
            HDL+WFQ++S+DVIGFLL CV   +F+ TK  LF C+K     K +K
Sbjct: 481 LHDLSWFQYHSLDVIGFLLLCVVGVVFIITKFCLFCCRKTANMGKKKK 528


>g207569 UDP glucuronosyltransferase-2
          Length = 529

 Score =  772 bits (1972), Expect = 0.0
 Identities = 361/528 (68%), Positives = 429/528 (80%), Gaps = 1/528 (0%)

Query: 1   MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA 60
           MSMK TS  LLIQL CYF  G+CGKVLVWPTE+SHW+NIK IL+EL QRGHEVTVL SSA
Sbjct: 1   MSMKQTSVFLLIQLICYFRPGACGKVLVWPTEYSHWINIKIILNELAQRGHEVTVLVSSA 60

Query: 61  SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMWTF 119
           SI  +P   S++ FE+Y V L+K++ E    + +  W  +    + W+Y+S++Q++   +
Sbjct: 61  SILIEPTKESSINFEIYSVPLSKSDLEYSFAKWIDEWTRDFETLSIWTYYSKMQKVFNEY 120

Query: 120 NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA 179
           +D++   CK ++ NK LMKKLQ S+FDV+LADAV P GELLAELLK P VYSLRF PGY
Sbjct: 121 SDVVENLCKALIWNKSLMKKLQGSQFDVILADAVGPCGELLAELLKTPLVYSLRFCPGYR 180

Query: 180 IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV 239
            EK SGGL  PPSYVPVV+SELSD+MTF+ERVKNM+ +LYF+FWFQ F  K W QFYS+V
Sbjct: 181 CEKFSGGLPLPPSYVPVVLSELSDRMTFVERVKNMLQMLYFDFWFQPFKEKSWSQFYSDV 240

Query: 240 LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS 299
           LGRPTTL+E M KADIWLIR +WD +FPHP LPN +FVGGLHCKPAKPLP+EMEEFVQSS
Sbjct: 241 LGRPTTLTEMMGKADIWLIRTFWDLEFPHPFLPNFDFVGGLHCKPAKPLPREMEEFVQSS 300

Query: 300 GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ 359
           GE+GVVVFSLGSMV N +EE+ANV+ASALA+IPQKV+WRFDG KPDTLG NTRLYKWIPQ
Sbjct: 301 GEHGVVVFSLGSMVKNLTEEKANVVASALAQIPQKVVWRFDGKKPDTLGSNTRLYKWIPQ 360

Query: 360 NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT 419
           NDLLGHPKTKAF+ HGG NGIYEAIYHG+P+VG+P+F DQ DNI HM AKGAAV ++F
Sbjct: 361 NDLLGHPKTKAFVAHGGTNGIYEAIYHGIPIVGIPLFADQPDNINHMVAKGAAVRVDFSI 420

Query: 420 MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA 479
           +++  LL AL+ V+ D SYKENAMRLSRIHHDQPVKPLDRAVFWIE+VMRHKGAKHLRS
Sbjct: 421 LSTTGLLTALKIVMNDPSYKENAMRLSRIHHDQPVKPLDRAVFWIEYVMRHKGAKHLRST 480

Query: 480 AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEK 527
            HDL+WFQ++S+DVIGFLL CV   +F+ TK  LF C+K     K +K
Sbjct: 481 LHDLSWFQYHSLDVIGFLLLCVVGVVFIITKFCLFCCRKTANMGKKKK 528
```

```
>g18146841 UGT2B21; UDP-glucuronosyltransferase 2B21; EC 2.7.1.17;
           UDP-glycosyltransferase
           Length = 528

 Score =  771 bits (1969), Expect = 0.0
 Identities = 361/528 (68%), Positives = 434/528 (81%), Gaps = 1/528 (0%)

Query: 3   MKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSASI 62
           MK   ALLL+QL C+F SGSCGKVLVWP EFSHWMNI+ IL+EL++RGHEVTVL  S  I
Sbjct: 1   MKRILALLLLQLCCHFHSGSCGKVLVWPMEFSHWMNIQVILEELIRRGHEVTVLRPSCFI 60

Query: 63  SFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPK-DTFWSYFSQVQEIMWTFND 121
             D N+ S +KFE +  S T+  +E I   LV  W      DT   YF +V+++   F+D
Sbjct: 61  FVDVNTTSEIKFETFHTSFTRDYWEKIFTDLVTTWLNTGSVDTCLDYFPEVEKLFKHFSD 120

Query: 122 ILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYAIE 181
               CK++VSNKK MK LQESRFD++LADAV P GEL+AE+L IPFVYSLRFSPG+  E
Sbjct: 121 EWENVCKELVSNKKFMKNLQESRFDILLADAVGPCGELVAEILHIPFVYSLRFSPGFQAE 180

Query: 182 KHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEVLG 241
           K +GGLL PPSYVPV+MS LS +MTF+ERVKNMI +LYF+FWF+ FD K+WD+ YSE+LG
Sbjct: 181 KRAGGLLLPPSYVPVIMSGLSGEMTFMERVKNMICMLYFDFWFETFDEKRWDKLYSEILG 240

Query: 242 RPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSSGE 301
           +P+TL ETM+KAD+WLIR+YWD +FPHP LPN +++GGLHCKPAKPLPKEMEEFVQSSGE
Sbjct: 241 KPSTLYETMSKADMWLIRSYWDMEFPHPSLPNFDYIGGLHCKPAKPLPKEMEEFVQSSGE 300

Query: 302 NGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQND 361
           +G+VVFSLGSM+ N ++E+AN+IASAL +IPQKVLWRFDG KPDTLG NTRLYKWIPQND
Sbjct: 301 HGIVVFSLGSMIRNMTDEKANLIASALGQIPQKVLWRFDGKKPDTLGANTRLYKWIPQND 360

Query: 362 LLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKTMT 421
           LLGHPKT+AFITHGG NGIYEAIYHG+PMVG+P+FG+Q DNIAHMKAKGAA+++ F +++
Sbjct: 361 LLGHPKTRAFITHGGANGIYEAIYHGIPMVGLPLFGEQYDNIAHMKAKGAAMKLEFNSLS 420

Query: 422 SEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSAAH 481
           S DLL AL+TVI + SYKENAM LS IHHDQP+KPLDRAVFWIE+VM+HKGAKHLR  AH
Sbjct: 421 STDLLNALKTVINNPSYKENAMWLSTIHHDQPMKPLDRAVFWIEYVMQHKGAKHLRPLAH 480

Query: 482 DLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 529
           +LTW+Q++S+DVIGFLL CVA   FL  KC LF  QKF +T K +KRE
Sbjct: 481 NLTWYQYHSLDVIGFLLACVAAITFLIIKCCLFCFQKFMETGKKKKRE 528


>g458395 UDP-glucuronosyltransferase; EC 2.4.1.17
           Length = 530

 Score =  755 bits (1929), Expect = 0.0
 Identities = 354/530 (66%), Positives = 427/530 (79%), Gaps = 1/530 (0%)

Query: 1   MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA 60
           MS KW SALLL+Q+S  F SG+CGKVLVWP E+SHWMNIK IL+ELVQ+GHEVTVL  SA
Sbjct: 1   MSGKWISALLLLQISFCFKSGNCGKVLVWPMEYSHWMNIKIILEELVQKGHEVTVLRPSA 60

Query: 61  SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMWTF 119
            +  DP   S LKF  +P S +  + E+   + V  W  ELP+DT   SYF  +Q+ +  +
Sbjct: 61  FVFLDPKETSDLKFVTFPTSFSSHDLENFFTRFVNVWTYELPRDTCLSYFLYLQDTIDEY 120

Query: 120 NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA 179
           +D     CK+ VSNK+ M KLQES+FDVV +DA+ P GEL+AELL+IPF+YSLRFSPGY
Sbjct: 121 SDYCLTVCKEAVSNKQFMTKLQESKFDVVFSDAIGPCGELIAELLQIPFLYSLRFSPGYT 180

Query: 180 IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV 239
           IE++ GG+LFPPSYVP++ S L+ QMTFIERV NMI +LYF+FWFQ F  KKWD FYS+
Sbjct: 181 IEQYIGGVLFPPSYVPMIFSGLAGQMTFIERVHNMICMLYFDFWFQTFREKKWDPFYSKT 240

Query: 240 LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS 299
```

```
           LGRPTTL+E M KA++WLIR+YWD +FPHP+ PNV+++GGLHCKPAKPLPK++E+FVQSS
Sbjct: 241 LGRPTTLAEIMGKAEMWLIRSYWDLEFPHPISPNVDYIGGLHCKPAKPLPKDIEDFVQSS 300

Query: 300 GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ 359
           GE+GVVVFSLGSMV N +EE+AN+IA ALA+IPQKVLWRFDG KP TLG NTRLYKW+PQ
Sbjct: 301 GEHGVVVFSLGSMVRNMTEEKANIIAWALAQIPQKVLWRFDGKKPPTLGPNTRLYKWLPQ 360

Query: 360 NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT 419
           NDLLGHPKTKAF+THGG NGIYEAI+HG+PM+G+P+F +Q DNIAHM AKGAAVE+NF+T
Sbjct: 361 NDLLGHPKTKAFVTHGGANGIYEAIHHGIPMIGIPLFAEQHDNIAHMVAKGAAVEVNFRT 420

Query: 420 MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA 479
           M+  DLL AL  VI +  YK+NAM LS IHHDQP KPLDRAVFWIEFVMRHKGAKHLRS
Sbjct: 421 MSKSDLLNALEEVIDNPFYKKNAMWLSTIHHDQPTKPLDRAVFWIEFVMRHKGAKHLRSL 480

Query: 480 AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 529
            H+L W+Q++S+DVIGFLL+CVA  + L  KCFLF  + F K  K K E
Sbjct: 481 GHNLPWYQYHSLDVIGFLLSCVAVTVVLALKCFLFVYRFFVKKEKKTKNE 530


>g458397 UDP-glucuronosyltransferase; EC 2.4.1.17
          Length = 530

 Score =  747 bits (1908), Expect = 0.0
 Identities = 351/530 (66%), Positives = 423/530 (79%), Gaps = 1/530 (0%)

Query: 1   MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA 60
           M  KW SALLL+Q+S  F SG+CGKVLVWP E+SHWMNIK IL+ELVQ+GHEVTVL  SA
Sbjct: 1   MPGKWISALLLLQISFCFKSGNCGKVLVWPMEYSHWMNIKIILEELVQKGHEVTVLRPSA 60

Query: 61  SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMWTF 119
           S+  DP   S LKF  +P S +  + E+   + V  W  ELP+DT  SYF  +Q+ +  +
Sbjct: 61  SVFLDPKETSHLKFVTFPTSFSSHDLENFFTRFVSVWTYELPRDTCLSYFLYLQDTIDEY 120

Query: 120 NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA 179
           +D     CK+ VSNK+ M KLQES+FDVV +DA+ P GEL+AELL+IPF+YSLRFSPGY
Sbjct: 121 SDYCLTVCKEAVSNKQFMTKLQESKFDVVFSDAIGPCGELIAELLQIPFLYSLRFSPGYT 180

Query: 180 IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV 239
           IEK+ GG+LFPPSYVP++ S L+ QMTFIERV NMI +LYF+FWFQ F  KKWD FYS+
Sbjct: 181 IEKYIGGVLFPPSYVPMIFSGLAGQMTFIERVHNMICMLYFDFWFQTFREKKWDPFYSKT 240

Query: 240 LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS 299
           LGRPTTL+E M KA++WLIR+YWD +FPHP+ PNV+++GGLHCKPAKPLPK++E+FVQSS
Sbjct: 241 LGRPTTLAEIMGKAEMWLIRSYWDLEFPHPISPNVDYIGGLHCKPAKPLPKDIEDFVQSS 300

Query: 300 GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ 359
           GE+GVVVFSLGSMV N +EE+AN+IA ALA+IPQKVLWRFDG KP TLG NTRLYKW+PQ
Sbjct: 301 GEHGVVVFSLGSMVRNMTEEKANIIAWALAQIPQKVLWRFDGKKPTTLGPNTRLYKWLPQ 360

Query: 360 NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT 419
           NDLLGHPKTKAF+THGG NGIYEAI+HG+PM+G+P+FG+Q DNIAHM AKGAA  +NF+T
Sbjct: 361 NDLLGHPKTKAFVTHGGANGIYEAIHHGIPMIGIPLFGEQHDNIAHMVAKGAAATVNFRT 420

Query: 420 MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA 479
           M+  DLL AL   I +  YK+NAM LS IHHDQP KPLDRAVFWIEFVMRHKGA HLRS
Sbjct: 421 MSKSDLLNALEEDIDNPFYKKNAMWLSTIHHDQPTKPLDRAVFWIEFVMRHKGALHLRSL 480

Query: 480 AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 529
            H+L W+ ++S+DVIGFLL+CVA  + L  KCFLF  + F K  K K E
Sbjct: 481 GHNLPWYLYHSLDVIGFLLSCVAVTVVLALKCFLFVYRFFVKKEKKTKNE 530


>g20381430 UDP-glucuronosyltransferase 2 family, member 5
          Length = 530

 Score =  747 bits (1907), Expect = 0.0
```

```
 Identities = 355/530 (66%), Positives = 421/530 (78%), Gaps = 1/530 (0%)

Query: 1   MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA 60
           M  KW SALLL+Q+SC F S  CGKVLVWP EFSHWMNIK ILDELVQRGHEVTVL  SA
Sbjct: 1   MPGKWISALLLLQISCCFRSVKCGKVLVWPMEFSHWMNIKIILDELVQRGHEVTVLRPSA 60

Query: 61  SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMWTF 119
               DP     LKFE +P S++K   E+   + V  W  E+P+DT  SY   +Q ++  F
Sbjct: 61  YYVLDPKKSPGLKFETFPTSVSKDNLENFFIKFVDVWTYEMPRDTCLSYSPLLQNMIDEF 120

Query: 120 NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA 179
           +D     CKD+VSNK+LM KLQES+FDV+L+D V   GEL+AELL+IPF+YS+RFSPGY
Sbjct: 121 SDYFLSLCKDVVSNKELMTKLQESKFDVLLSDPVASCGELIAELLQIPFLYSIRFSPGYQ 180

Query: 180 IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV 239
           IEK SG  L PPSYVPV++S L  QMTFIER+KNMI +LYF+FWFQ+F+ KKWD FYSE
Sbjct: 181 IEKSSGRFLLPPSYVPVILSGLGGQMTFIERIKNMICMLYFDFWFQMFNDKKWDSFYSEY 240

Query: 240 LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS 299
           LGRPTTL ETM +A++WLIR+ WD +FPHP LPNV++VGGLHCKPAKPLPK+MEEFVQSS
Sbjct: 241 LGRPTTLVETMGQAEMWLIRSNWDLEFPHPTLPNVDYVGGLHCKPAKPLPKDMEEFVQSS 300

Query: 300 GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ 359
           G++GVVVFSLGSMVSN +EE+AN IA ALA+IPQKVLW+FDG  P TLG NTR+YKW+PQ
Sbjct: 301 GDHGVVVFSLGSMVSNMTEEKANAIAWALAQIPQKVLWKFDGKTPATLGHNTRVYKWLPQ 360

Query: 360 NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT 419
           NDLLGHPKTKAF+THGG NG+YEAIYHG+PM+G+P+FG+Q DNIAHM AKGAAV +N +T
Sbjct: 361 NDLLGHPKTKAFVTHGGANGVYEAIYHGIPMIGIPLFGEQHDNIAHMVAKGAAVALNIRT 420

Query: 420 MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA 479
           M+  D+L AL  VI +  YK+NAM LS IHHDQP+KPLDRAVFWIEFVMRHK AKHLR
Sbjct: 421 MSKSDVLNALEEVIENPFYKKNAMWLSTIHHDQPMKPLDRAVFWIEFVMRHKRAKHLRPL 480

Query: 480 AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 529
            H+LTW+Q++S+DVIGFLL+CVAT I L  KC LF  + F K     K E
Sbjct: 481 GHNLTWYQYHSLDVIGFLLSCVATTIVLSVKCLLFIYRFFVKKENKMKNE 530


>g55120 unnamed protein product; UDP-glucuronosyltransferase
          precursor (530 AA)
          Length = 530

 Score =  745 bits (1903), Expect = 0.0
 Identities = 354/530 (66%), Positives = 421/530 (78%), Gaps = 1/530 (0%)

Query: 1   MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA 60
           M  KW SALLL+Q+SC F S  CGKVLVWP EFSHWMNIK ILDELVQRGHEVTVL  SA
Sbjct: 1   MPGKWISALLLLQISCCFRSVKCGKVLVWPMEFSHWMNIKIILDELVQRGHEVTVLRPSA 60

Query: 61  SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMWTF 119
               DP     LKFE +P S++K   E+   + V  W  E+P+DT  SY   +Q ++  F
Sbjct: 61  YYVLDPKKSPGLKFETFPTSVSKDNLENFFIKFVDVWTYEMPRDTCLSYSPLLQNMIDEF 120

Query: 120 NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA 179
           +D     CKD+VSNK+LM KLQES+FDV+L+D V   GEL+AELL+IPF+YS+RFSPGY
Sbjct: 121 SDYFLSLCKDVVSNKELMTKLQESKFDVLLSDPVASCGELIAELLQIPFLYSIRFSPGYQ 180

Query: 180 IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV 239
           IEK SG  L PPSYVPV++S L  QMTFIER+KNMI +LYF+FWFQ+F+ KKWD FYSE
Sbjct: 181 IEKSSGRFLLPPSYVPVILSGLGGQMTFIERIKNMICMLYFDFWFQMFNDKKWDSFYSEY 240

Query: 240 LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS 299
           LGRPTTL ETM +A++WLIR+ WD +FPHP LPNV++VGGLHCKPAKPLPK+MEEFVQSS
Sbjct: 241 LGRPTTLVETMGQAEMWLIRSNWDLEFPHPTLPNVDYVGGLHCKPAKPLPKDMEEFVQSS 300

Query: 300 GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ 359
```

```
          G++GVVVFSLGSMVSN +EE+AN IA ALA+IPQKVLW+FDG  P TLG NTR+YKW+PQ
Sbjct: 301 GDHGVVVFSLGSMVSNMTEEKANAIAWALAQIPQKVLWKFDGKTPATLGHNTRVYKWLPQ 360

Query: 360 NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT 419
           NDLLGHPKTKAF+THGG NG+YEAIYHG+PM+G+P+FG+Q DNIAHM AKGAAV +N +T
Sbjct: 361 NDLLGHPKTKAFVTHGGANGVYEAIYHGIPMIGIPLFGEQHDNIAHMVAKGAAVALNIRT 420

Query: 420 MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA 479
           M+  D+L AL  VI +  YK+NA+ LS IHHDQP+KPLDRAVFWIEFVMRHK AKHLR
Sbjct: 421 MSKSDVLNALEEVIENPFYKKNAIWLSTIHHDQPMKPLDRAVFWIEFVMRHKRAKHLRPL 480

Query: 480 AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 529
            H+LTW+Q++S+DVIGFLL+CVAT I L  KC LF  + F K     K E
Sbjct: 481 GHNLTWYQYHSLDVIGFLLSCVATTIVLSVKCLLFIYRFFVKKENKMKNE 530


>g15929692 RIKEN cDNA 9430041C03
          Length = 530

 Score =  742 bits (1894), Expect = 0.0
 Identities = 354/530 (66%), Positives = 418/530 (78%), Gaps = 1/530 (0%)

Query: 1   MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA 60
           M  KW SALLL+Q+SC F S  CGKVLVWP EFSHWMNIK ILDELVQRGHEVTVL  SA
Sbjct: 1   MPGKWISALLLLQISCCFRSVKCGKVLVWPMEFSHWMNIKIILDELVQRGHEVTVLRPSA 60

Query: 61  SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMWTF 119
               DP     LKFE +P S++K   E+   + V  W  E+P+DT  SY   +Q ++  F
Sbjct: 61  YYVLDPKKSPGLKFETFPTSVSKDNLENFFIKFVDVWTYEMPRDTCLSYSPLLQNMIDEF 120

Query: 120 NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA 179
           +D     CKD+VSNK+LM KLQES+FDV+L+D V   GEL+AELL+IPF+YS+RFSPGY
Sbjct: 121 SDYFLSLCKDVVSNKELMTKLQESKFDVLLSDPVASCGELIAELLQIPFLYSIRFSPGYQ 180

Query: 180 IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV 239
           IEK SG  L PPSYVPV++S L  QMTFIERVKNMI  LYF+FWFQ+F+ KKWD FYSE
Sbjct: 181 IEKSSGRFLLPPSYVPVILSGLGGQMTFIERVKNMICRLYFDFWFQMFNDKKWDSFYSEY 240

Query: 240 LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS 299
           LGRPTTL+ETM KA++WLIR+ WD +FPHP LPNV++VGGLHCKPAKPLPK+MEEFVQSS
Sbjct: 241 LGRPTTLAETMGKAEMWLIRSNWDLEFPHPTLPNVDYVGGLHCKPAKPLPKDMEEFVQSS 300

Query: 300 GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ 359
           G++GVVVFSLGSMVSN +EE+AN IA ALA+IPQKVLW+FDG  P TLG NTR+YKW+PQ
Sbjct: 301 GDHGVVVFSLGSMVSNMTEEKANTIAWALAQIPQKVLWKFDGKTPATLGHNTRVYKWLPQ 360

Query: 360 NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT 419
           NDLLGHPKTKAF+THGG NG+YE IYHG+PM+G+P+FG+Q DNIAHM AKGAAV +N +T
Sbjct: 361 NDLLGHPKTKAFVTHGGANGVYEVIYHGIPMIGIPLFGEQHDNIAHMVAKGAAVTLNIRT 420

Query: 420 MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA 479
           M+  D+L AL  VI +  YK+NA+ LS IHHDQP KPLDRAVFW+EFVMRHK AKHLRS
Sbjct: 421 MSRSDVLNALEEVIDNPFYKKNAIWLSTIHHDQPTKPLDRAVFWVEFVMRHKRAKHLRSL 480

Query: 480 AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE 529
            H+LTW Q++ +DVIGFLL+CVA  I L  KC LF  + F K  K  K E
Sbjct: 481 GHNLTWHQYHFLDVIGFLLSCVAVTIVLTVKCLLFIYRFFVKKEKKIKNE 530


  Database: gb137rodp
    Posted date:  Sep 11, 2003 11:24 AM
  Number of letters in database: 25,169,402
  Number of sequences in database:  74,095

Lambda     K      H
   0.324    0.138    0.428
```

```
Gapped
Lambda      K       H
   0.270   0.0470    0.230


Matrix: BLOSUM62
Gap Penalties: Existence: 11, Extension: 1
Number of Hits to DB: 27352210
Number of Sequences: 74095
Number of extensions: 1162215
Number of successful extensions: 2617
Number of sequences better than 10.0: 77
Number of HSP's better than 10.0 without gapping: 72
Number of HSP's successfully gapped in prelim test: 5
Number of HSP's that attempted gapping in prelim test: 2394
Number of HSP's gapped (non-prelim): 84
length of query: 529
length of database: 25,169,402
effective HSP length: 49
effective length of query: 480
effective length of database: 21,538,747
effective search space: 10338598560
effective search space used: 10338598560
T: 11
A: 40
X1: 15 ( 7.0 bits)
X2: 38 (14.8 bits)
X3: 64 (24.9 bits)
S1: 41 (22.0 bits)
```

Clear Query Window

Submit sequences to: BLAST2 Submit Reset

IncyteGenomics

Exhibit C
with Response dated 4/26/04
In USSN: 09/980,729

1: AAA42313. UDP-glucuronosylt...[gi:207581]

```
LOCUS       AAA42313                529 aa            linear   ROD 27-APR-1993
DEFINITION  UDP-glucuronosyltransferase (EC 2.4.1.17).
ACCESSION   AAA42313
VERSION     AAA42313.1  GI:207581
DBSOURCE    locus RATUDPGTP accession M13506.1
KEYWORDS    .
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (residues 1 to 529)
  AUTHORS   Mackenzie,P.I.
  TITLE     Rat liver UDP-glucuronosyltransferase. Sequence and expression of a
            cDNA encoding a phenobarbital-inducible form
  JOURNAL   J. Biol. Chem. 261 (13), 6119-6125 (1986)
  MEDLINE   86196018
   PUBMED   3084479
COMMENT     Method: conceptual translation.
FEATURES             Location/Qualifiers
     source          1..529
                     /organism="Rattus norvegicus"
                     /db_xref="taxon:10116"
     Protein         1..529
                     /name="UDP-glucuronosyltransferase (EC 2.4.1.17)"
     sig_peptide     1..24
                     /note="UDP-glucuronosyltransferase, signal peptide"
     mat_peptide     25..529
                     /product="UDP-glucuronosyltransferase"
     CDS             1..529
                     /coded_by="M13506.1:26..1615"
ORIGIN
        1 msmkqtsvfl liqlicyfrp gacgkvlvwp teyshwinik iilnelaqrg hevtvlvssa
       61 silieptkes sinfeiysvp lsksdleysf akwidewtrd fetlsiwtyy skmqkvfney
      121 sdvvenlcka liwnkslmkk lqgsqfdvil adavgpcgel laellktplv yslrfcpgyr
      181 cekfsgglpl ppsyvpvvls elsdrmtfve rvknmlqmly fdfwfqpfke kswsqfysdv
      241 lgrpttltem mgkadiwlir tfwdlefphp flpnfdfvgg lhckpakplp remeefvqss
      301 gehgvvvfsl gsmvknltee kanvvasala qipqkvvwrf dgkkpdtlgs ntrlykwipq
      361 ndllghpktk afvahggtng iyeaiyhgip ivgiplfadq pdninhmvak gaavrvdfsi
      421 lsttglltal kivmndpsyk enamrlsrih hdqpvkpldr avfwieyvmr hkgakhlrst
      481 lhdlswfqyh sldvigflll cvvgvvfiit kfclfccrkt anmgkkkke
//
```

SeqServer®
biology in silico

ClustalW Results

Sequences | Help

Retrieval | BLAST2 | FASTA | ClustalW | GCG Assembly | Phrap | Translation

 SeqServer Version 4.6 Jan 2002

- 2912330CD1
- g207581

CLUSTAL W (1.7) Multiple Sequence Alignments

```
Sequence format is Pearson
Sequence 1: 2912330CD1      529 aa
Sequence 2: g207581         529 aa
Start of Pairwise alignments
Aligning...
Sequences (1:2) Aligned. Score:  68
Start of Multiple Alignment
There are 1 groups
Aligning...
Group 1: Sequences:   2      Score:6064
Alignment Score 2397
CLUSTAL-Alignment file created  [baa4vairG.aln]
CLUSTAL W (1.7) multiple sequence alignment


2912330CD1      MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA
g207581         MSMKQTSVFLLIQLICYFRPGACGKVLVWPTEYSHWINIKIILNELAQRGHEVTVLVSSA
                **** **.:***** *** .*:**********:***:*** **:**.*********.***

2912330CD1      SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMWTF
g207581         SILIEPTKESSINFEIYSVPLSKSDLEYSFAKWIDEWTRDFETLSIWTYYSKMQKVFNEY
                ** ::*.. *:::**:*.*.*:*:::*  : : :..*: :: . ::*:*:*::*:::  :

2912330CD1      NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA
g207581         SDVVENLCKALIWNKSLMKKLQGSQFDVILADAVGPCGELLAELLKTPLVYSLRFCPGYR
                .*::.::** :: **.****** *:***:***** * ********* *:******.***

2912330CD1      IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV
g207581         CEKFSGGLPLPPSYVPVVLSELSDRMTFVERVKNMLQMLYFDFWFQPFKEKSWSQFYSDV
                 **.**** :********:*****:***:******: :***:**** *. *.*.****:*

2912330CD1      LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS
g207581         LGRPTTLTEMMGKADIWLIRTFWDLEFPHPFLPNFDFVGGLHCKPAKPLPREMEEFVQSS
                *******:* *.********.:**::****:***.:**************:*********

2912330CD1      GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ
g207581         GEHGVVVFSLGSMVKNLTEEKANVVASALAQIPQKVVWRFDGKKPDTLGSNTRLYKWIPQ
```

```
                **:***********.* :**:***:*****:*****:*****:****** **********

2912330CD1      NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT
g207581         NDLLGHPKTKAFVAHGGTNGIYEAIYHGIPIVGIPLFADQPDNINHMVAKGAAVRVDFSI
                ***********::*** **********:*:**:*:*.** *** ** ******.::*.

2912330CD1      MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA
g207581         LSTTGLLTALKIVMNDPSYKENAMRLSRIHHDQPVKPLDRAVFWIEYVMRHKGAKHLRST
                ::: .** **: *:.*.*****************************:***********:

2912330CD1      AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE
g207581         LHDLSWFQYHSLDVIGFLLLCVVGVVFIITKFCLFCCRKTANMGKK-KKE
                 ***:***::*:******* **. .:*::**  **.*:*  :  *  *:*
```

Submit sequences to: BLAST2 Submit Reset

IncyteGenomics

1: NP_033493. UDP-glucuronosylt...[gi:6678501] BLink, Domains, Links

```
LOCUS       NP_033493                530 aa            linear   ROD 24-DEC-2003
DEFINITION  UDP-glucuronosyltransferase 2 family, member 5 [Mus musculus].
ACCESSION   NP_033493
VERSION     NP_033493.1  GI:6678501
DBSOURCE    REFSEQ: accession NM_009467.1
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (residues 1 to 530)
  AUTHORS   Kimura,T. and Owens,I.S.
  TITLE     Mouse UDP glucuronosyltransferase. cDNA and complete amino acid
            sequence and regulation
  JOURNAL   Eur. J. Biochem. 168 (3), 515-521 (1987)
   PUBMED   3117546
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from X06358.1.
FEATURES             Location/Qualifiers
     source          1..530
                     /organism="Mus musculus"
                     /strain="C57Bl/6N"
                     /db_xref="taxon:10090"
                     /chromosome="5"
                     /map="5"
     Protein         1..530
                     /product="UDP-glucuronosyltransferase 2 family, member 5"
     variation       19
                     /allele="R"
                     /allele="S"
                     /db_xref="dbSNP:8258200"
     Region          24..528
                     /region_name="UDP-glucoronosyl and UDP-glucosyl
                     transferase"
                     /note="UDPGT"
                     /db_xref="CDD:22944"
     variation       253
                     /allele="Q"
                     /allele="K"
                     /db_xref="dbSNP:8258202"
     CDS             1..530
                     /gene="Ugt2b5"
                     /coded_by="NM_009467.1:13..1605"
                     /note="go_component: microsome [goid 0005792] [evidence
                     IEA];
                     go_component: integral to membrane [goid 0016021]
                     [evidence IEA];
                     go_function: transferase activity, transferring glycosyl
                     groups [goid 0016757] [evidence IEA];
                     go_function: glucuronosyltransferase activity [goid
                     0015020] [evidence IEA];
                     go_function: transferase activity [goid 0016740] [evidence
                     IEA];
                     go_function: transferase activity, transferring hexosyl
                     groups [goid 0016758] [evidence IEA];
                     go_process: metabolism [goid 0008152] [evidence IEA]"
                     /db_xref="GeneID:22238"
                     /db_xref="LocusID:22238"
                     /db_xref="MGI:98900"
ORIGIN
        1 mpgkwisall llqisccfrs vkcgkvlvwp mefshwmnik iildelvqrg hevtvlrpsa
       61 yyvldpkksp glkfetfpts vskdnlenff ikfvdvwtye mprdtclsys pllqnmidef
```

```
      121 sdyflslckd vvsnkelmtk lqeskfdvll sdpvascgel iaellqipfl ysirfspgyq
      181 iekssgrfll ppsyvpvils glggqmtfie riknmicmly fdfwfqmfnd kkwdsfysey
      241 lgrpttlvet mgqaemwlir snwdlefphp tlpnvdyvgg lhckpakplp kdmeefvqss
      301 gdhgvvvfsl gsmvsnmtee kanaiawala qipqkvlwkf dgktpatlgh ntrvykwlpq
      361 ndllghpktk afvthggang vyeaiyhgip migiplfgeq hdniahmvak gaavalnirt
      421 msksdvlnal eevienpfyk knaiwlstih hdqpmkpldr avfwiefvmr hkrakhlrpl
      481 ghnltwyqyh sldvigflls cvattivlsv kcllfiyrff vkkenkmkne
//
```

SeqServer®
biology *in silico*

ClustalW Results

Sequences | Help

Retrieval | BLAST2 | FASTA | ClustalW | GCG Assembly | Phrap | Translation

- 2912330CD1
- g6678501

CLUSTAL W (1.7) Multiple Sequence Alignments

```
Sequence format is Pearson
Sequence 1: 2912330CD1      529 aa
Sequence 2: g6678501        530 aa
Start of Pairwise alignments
Aligning...
Sequences (1:2) Aligned. Score:  66
Start of Multiple Alignment
There are 1 groups
Aligning...
Group 1: Sequences:   2      Score:5979
Alignment Score 2326
CLUSTAL-Alignment file created  [baaoraiyA.aln]
CLUSTAL W (1.7) multiple sequence alignment


2912330CD1      MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA
g6678501        MPGKWISALLLLQISCCFRSVKCGKVLVWPMEFSHWMNIKIILDELVQRGHEVTVLRPSA
                *. ** *****:*:** * * .******** ********* *************** .**

2912330CD1      SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMWTF
g6678501        YYVLDPKKSPGLKFETFPTSVSKDNLENFFIKFVDVWTYEMPRDTCLSYSPLLQNMIDEF
                   :**:... ****.:*.*::* ::*::: ::*. *: *:*:**  ** . :*:::  *

2912330CD1      NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA
g6678501        SDYFLSLCKDVVSNKELMTKLQESKFDVLLSDPVASCGELIAELLQIPFLYSIRFSPGYQ
                .* : .:***:****:**.*****:***:*:*.* . ***:****:***:**:******

2912330CD1      IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV
g6678501        IEKSSGRFLLPPSYVPVILSGLGGQMTFIERIKNMICMLYFDFWFQMFNDKKWDSFYSEY
                *** ** :*:********::* *..*******:**** :***:****:*: ****.****

2912330CD1      LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS
g6678501        LGRPTTLVETMGQAEMWLIRSNWDLEFPHPTLPNVDYVGGLHCKPAKPLPKDMEEFVQSS
                ******* ***.:*::****. **::**** ****::**************:********

2912330CD1      GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ
g6678501        GDHGVVVFSLGSMVSNMTEEKANAIAWALAQIPQKVLWKFDGKTPATLGHNTRVYKWLPQ
```

```
                *::************* :**:**.** ***:*******:***:.* *** ***:***:**

2912330CD1      NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT
g6678501        NDLLGHPKTKAFVTHGGANGVYEAIYHGIPMIGIPLFGEQHDNIAHMVAKGAAVALNIRT
                ************:**** **:*******:**:*:*:**:* ****** ****** :*::*

2912330CD1      MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA
g6678501        MSKSDVLNALEEVIENPFYKKNAIWLSTIHHDQPMKPLDRAVFWIEFVMRHKRAKHLRPL
                *:..*:*.**. ** :. **:**: ** ******:**************** *****.

2912330CD1      AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE
g6678501        GHNLTWYQYHSLDVIGFLLSCVATTIVLSVKCLLFIYRFFVKKENKMKNE
                .*:***:*::*:*******:****:*.* .**:**  : * *..:  *.*
```

Submit sequences to: BLAST2 Submit Reset

IncyteGenomics

1: CAA29657. unnamed protein p...[gi:55120] BLink, Domains, Links

```
LOCUS       CAA29657                530 aa            linear   ROD 12-SEP-1993
DEFINITION  unnamed protein product [Mus musculus].
ACCESSION   CAA29657
VERSION     CAA29657.1  GI:55120
DBSOURCE    embl locus MMUDPGT, accession X06358.1
KEYWORDS    .
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (residues 1 to 530)
  AUTHORS   Kimura,T. and Owens,I.S.
  TITLE     Mouse UDP glucuronosyltransferase. cDNA and complete amino acid
            sequence and regulation
  JOURNAL   Eur. J. Biochem. 168 (3), 515-521 (1987)
  MEDLINE   88029469
   PUBMED   3117546
COMMENT     Data kindly reviewed (07-SEP-1988) by OWENS I.S.
FEATURES             Location/Qualifiers
     source          1..530
                     /organism="Mus musculus"
                     /strain="C57BL/6N."
                     /db_xref="taxon:10090"
                     /clone="UDPGTm-1"
                     /clone_lib="lambda gt11"
     Protein         1..530
                     /name="unnamed protein product"
     CDS             1..530
                     /coded_by="X06358.1:13..1605"
                     /note="UDP-glucuronosyltransferase precursor (530 AA)"
                     /db_xref="GOA:P17717"
                     /db_xref="Swiss-Prot:P17717"
ORIGIN
        1 mpgkwisall llqisccfrs vkcgkvlvwp mefshwmnik iildelvqrg hevtvlrpsa
       61 yyvldpkksp glkfetfpts vskdnlenff ikfvdvwtye mprdtclsys pllqnmidef
      121 sdyflslckd vvsnkelmtk lqeskfdvll sdpvascgel iaellqipfl ysirfspgyq
      181 iekssgrfll ppsyvpvils glggqmtfie riknmicmly fdfwfqmfnd kkwdsfysey
      241 lgrpttlvet mgqaemwlir snwdlefphp tlpnvdyvgg lhckpakplp kdmeefvqss
      301 gdhgvvvfsl gsmvsnmtee kanaiawala qipqkvlwkf dgktpatlgh ntrvykwlpq
      361 ndllghpktk afvthggang vyeaiyhgip migiplfgeq hdniahmvak gaavalnirt
      421 msksdvlnal eevienpfyk knaiwlstih hdqpmkpldr avfwiefvmr hkrakhlrpl
      481 ghnltwyqyh sldvigflls cvattivlsv kcllfiyrff vkkenkmkne
//
```

SeqServer®
biology *in silico*

ClustalW Results

Sequences | Help

Retrieval | BLAST2 | FASTA | ClustalW | GCG Assembly | Phrap | Translation

 SeqServer Version 4.6 Jan 2002

- 2912330CD1
- g55120

CLUSTAL W (1.7) Multiple Sequence Alignments

```
Sequence format is Pearson
Sequence 1: 2912330CD1      529 aa
Sequence 2: g55120          530 aa
Start of Pairwise alignments
Aligning...
Sequences (1:2) Aligned. Score:  66
Start of Multiple Alignment
There are 1 groups
Aligning...
Group 1: Sequences:   2      Score:5979
Alignment Score 2326
CLUSTAL-Alignment file created  [baasfaqUD.aln]
CLUSTAL W (1.7) multiple sequence alignment

2912330CD1      MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA
g55120          MPGKWISALLLLQISCCFRSVKCGKVLVWPMEFSHWMNIKIILDELVQRGHEVTVLRPSA
                *. ** *****:*:** * * .******** ********* *************** .**

2912330CD1      SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMWTF
g55120          YYVLDPKKSPGLKFETFPTSVSKDNLENFFIKFVDVWTYEMPRDTCLSYSPLLQNMIDEF
                   :**:... ****.:*.*::* ::*::: ::*. *: *:*:**  ** . :*:::  *

2912330CD1      NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA
g55120          SDYFLSLCKDVVSNKELMTKLQESKFDVLLSDPVASCGELIAELLQIPFLYSIRFSPGYQ
                .* : .:***:****:**.*****:***:*:*.* . ***:****:***:**:******

2912330CD1      IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV
g55120          IEKSSGRFLLPPSYVPVILSGLGGQMTFIERIKNMICMLYFDFWFQMFNDKKWDSFYSEY
                *** ** :*:*******::* *..*******:**** :***:****:*: ****.****

2912330CD1      LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS
g55120          LGRPTTLVETMGQAEMWLIRSNWDLEFPHPTLPNVDYVGGLHCKPAKPLPKDMEEFVQSS
                ******* ***.:*::****. **::**** ****::**************:********

2912330CD1      GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ
g55120          GDHGVVVFSLGSMVSNMTEEKANAIAWALAQIPQKVLWKFDGKTPATLGHNTRVYKWLPQ
```

```
                *::************** :**:**.** ***:*******:***:.* *** ***:***:**

2912330CD1      NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT
g55120          NDLLGHPKTKAFVTHGGANGVYEAIYHGIPMIGIPLFGEQHDNIAHMVAKGAAVALNIRT
                ************:**** **:*******:**:*:*:**:* ****** ****** :*::*

2912330CD1      MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA
g55120          MSKSDVLNALEEVIENPFYKKNAIWLSTIHHDQPMKPLDRAVFWIEFVMRHKRAKHLRPL
                *:..*:*.**. ** :. **:**: ** ******:****************** *****.

2912330CD1      AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE
g55120          GHNLTWYQYHSLDVIGFLLSCVATTIVLSVKCLLFIYRFFVKKENKMKNE
                .*:***:*::*:*******:****:*.* .**:**  : * *..:  *.*
```

Submit sequences to: BLAST2 Submit Reset

IncyteGenomics

Exhibit D
with Response dated 4/26/04
In USSN: 09/980,729

1: P17717. UDP-glucuronosylt...[gi:136725]


```
LOCUS       P17717                  530 aa            linear   ROD 16-OCT-2001
DEFINITION  UDP-glucuronosyltransferase 2B5 precursor, microsomal (UDPGT)
            (M-1).
ACCESSION   P17717
VERSION     P17717  GI:136725
DBSOURCE    swissprot: locus UDB5_MOUSE, accession P17717;
            class: standard.
            created: Aug 1, 1990.
            sequence updated: Aug 1, 1990.
            annotation updated: Oct 16, 2001.
            xrefs: gi: 55119, gi: 55120, gi: 90521
            xrefs (non-sequence databases): MGI98900, InterProIPR002213,
            PfamPF00201, PROSITEPS00375
KEYWORDS    Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
            Signal; Multigene family; Microsome.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (residues 1 to 530)
  AUTHORS   Kimura,T. and Owens,I.S.
  TITLE     Mouse UDP glucuronosyltransferase. cDNA and complete amino acid
            sequence and regulation
  JOURNAL   Eur. J. Biochem. 168 (3), 515-521 (1987)
  MEDLINE   88029469
   PUBMED   3117546
  REMARK    SEQUENCE FROM N.A.
            STRAIN=C57BL/6N; TISSUE=Liver
COMMENT     ---------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------.
            [FUNCTION] UDPGT IS OF MAJOR IMPORTANCE IN THE CONJUGATION AND
            SUBSEQUENT ELIMINATION OF POTENTIALLY TOXIC XENOBIOTICS AND
            ENDOGENOUS COMPOUNDS.
            [CATALYTIC ACTIVITY] UDP-GLUCURONATE + ACCEPTOR = UDP + ACCEPTOR
            BETA-D-GLUCURONOSIDE.
            [SUBCELLULAR LOCATION] MICROSOMAL.
            [SIMILARITY] BELONGS TO THE UDP-GLYCOSYLTRANSFERASE FAMILY.
FEATURES             Location/Qualifiers
     source          1..530
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
     gene            1..530
                     /gene="UGT2B5"
     Protein         1..530
                     /gene="UGT2B5"
                     /product="UDP-glucuronosyltransferase 2B5 precursor,
                     microsomal"
                     /EC_number="2.4.1.17"
     Region          1..23
                     /gene="UGT2B5"
                     /region_name="Signal"
                     /note="BY SIMILARITY."
     Region          24..530
                     /gene="UGT2B5"
                     /region_name="Mature chain"
                     /note="UDP-GLUCURONOSYLTRANSFERASE 2B5."
     Site            316
```

```
                     /gene="UGT2B5"
                     /site_type="glycosylation"
                     /note="N-LINKED (GLCNAC...) (POTENTIAL)."
     Site            483
                     /gene="UGT2B5"
                     /site_type="glycosylation"
                     /note="N-LINKED (GLCNAC...) (POTENTIAL)."
     Region          494..510
                     /gene="UGT2B5"
                     /region_name="Transmembrane region"
                     /note="POTENTIAL."
ORIGIN
        1 mpgkwisall llqisccfrs vkcgkvlvwp mefshwmnik iildelvqrg hevtvlrpsa
       61 yyvldpkksp glkfetfpts vskdnlenff ikfvdvwtye mprdtclsys pllqnmidef
      121 sdyflslckd vvsnkelmtk lqeskfdvll sdpvascgel iaellqipfl ysirfspgyq
      181 iekssgrfll ppsyvpvils glggqmtfie riknmicmly fdfwfqmfnd kkwdsfysey
      241 lgrpttlvet mgqaemwlir snwdlefphp tlpnvdyvgg lhckpakplp kdmeefvqss
      301 gdhgvvvfsl gsmvsnmtee kanaiawala qipqkvlwkf dgktpatlgh ntrvykwlpq
      361 ndllghpktk afvthggang vyeaiyhgip migiplfgeq hdniahmvak gaavalnirt
      421 msksdvlnal eevienpfyk knaiwlstih hdqpmkpldr avfwiefvmr hkrakhlrpl
      481 ghnltwyqyh sldvigflls cvattivlsv kcllfiyrff vkkenkmkne
//
```

SeqServer®
biology *in silico*

ClustalW Results

Sequences | Help

Retrieval | BLAST2 | FASTA | ClustalW | GCG Assembly | Phrap | Translation

 SeqServer Version 4.6 Jan 2002

- 2912330CD1
- g136725

CLUSTAL W (1.7) Multiple Sequence Alignments

```
Sequence format is Pearson
Sequence 1: 2912330CD1      529 aa
Sequence 2: g136725         530 aa
Start of Pairwise alignments
Aligning...
Sequences (1:2) Aligned. Score:  66
Start of Multiple Alignment
There are 1 groups
Aligning...
Group 1: Sequences:   2      Score:5979
Alignment Score 2326
CLUSTAL-Alignment file created  [baac9ayTA.aln]
CLUSTAL W (1.7) multiple sequence alignment


2912330CD1      MSMKWTSALLLIQLSCYFSSGSCGKVLVWPTEFSHWMNIKTILDELVQRGHEVTVLASSA
g136725         MPGKWISALLLLQISCCFRSVKCGKVLVWPMEFSHWMNIKIILDELVQRGHEVTVLRPSA
                *. ** *****:*:** * * .******** ********* *************** .**

2912330CD1      SISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWA-ELPKDTFWSYFSQVQEIMWTF
g136725         YYVLDPKKSPGLKFETFPTSVSKDNLENFFIKFVDVWTYEMPRDTCLSYSPLLQNMIDEF
                   :**:... ****.:*.*::* ::*::: ::*. *: *:*:**  ** . :*:::  *

2912330CD1      NDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFSPGYA
g136725         SDYFLSLCKDVVSNKELMTKLQESKFDVLLSDPVASCGELIAELLQIPFLYSIRFSPGYQ
                .* : .:***:****:**.*****:***:*:*.* . ***:****:***:**:******

2912330CD1      IEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQFYSEV
g136725         IEKSSGRFLLPPSYVPVILSGLGGQMTFIERIKNMICMLYFDFWFQMFNDKKWDSFYSEY
                *** ** :*:*******::* *..*******:**** :***:****:*: ****.****

2912330CD1      LGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEFVQSS
g136725         LGRPTTLVETMGQAEMWLIRSNWDLEFPHPTLPNVDYVGGLHCKPAKPLPKDMEEFVQSS
                ******* ***.:*::****. **::**** ****::**************:********

2912330CD1      GENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYKWIPQ
g136725         GDHGVVVFSLGSMVSNMTEEKANAIAWALAQIPQKVLWKFDGKTPATLGHNTRVYKWLPQ
```

```
                *::************* :**:**.** ***:*******:***:.* *** ***:***:**

2912330CD1      NDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEINFKT
g136725         NDLLGHPKTKAFVTHGGANGVYEAIYHGIPMIGIPLFGEQHDNIAHMVAKGAAVALNIRT
                ************:**** **:*******:**:*:*:**:* ****** ****** :*::*

2912330CD1      MTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKHLRSA
g136725         MSKSDVLNALEEVIENPFYKKNAIWLSTIHHDQPMKPLDRAVFWIEFVMRHKRAKHLRPL
                *:..*:*.**. ** :. **:**: ** ******:**************** *****.

2912330CD1      AHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTRKIEKRE
g136725         GHNLTWYQYHSLDVIGFLLSCVATTIVLSVKCLLFIYRFFVKKENKMKNE
                .*:***:*::*:*******:****:*.* .**:**  : * *..:  *.*
```

Submit sequences to: BLAST2 Submit Reset

IncyteGenomics

SeqServer®
biology *in silico*

ClustalW Results

Sequences | Help

Retrieval | BLAST2 | FASTA | ClustalW | GCG Assembly | Phrap | Translation

 SeqServer Version 4.6 Jan 2002

- 2912330CD1
- g136729
- g31377618
- g29789078
- g13487900
- g10863941
- g8850236
- g6005930
- g4507815
- g41282213
- g136731

CLUSTAL W (1.7) Multiple Sequence Alignments

```
Sequence format is Pearson
Sequence 1: 2912330CD1      529 aa
Sequence 2: g136729         533 aa
Sequence 3: g31377618       530 aa
Sequence 4: g29789078       530 aa
Sequence 5: g13487900       534 aa
Sequence 6: g10863941       528 aa
Sequence 7: g8850236        533 aa
Sequence 8: g6005930        534 aa
Sequence 9: g4507815        531 aa
Sequence 10: g41282213       530 aa
Sequence 11: g136731         534 aa
Start of Pairwise alignments
Aligning...
Sequences (1:2) Aligned. Score:  41
Sequences (1:3) Aligned. Score:  41
Sequences (1:4) Aligned. Score:  41
Sequences (1:5) Aligned. Score:  41
Sequences (1:6) Aligned. Score:  90
Sequences (1:7) Aligned. Score:  41
Sequences (1:8) Aligned. Score:  41
Sequences (1:9) Aligned. Score:  38
Sequences (1:10) Aligned. Score:  41
Sequences (1:11) Aligned. Score:  41
Sequences (2:3) Aligned. Score:  65
```

```
Sequences (2:4) Aligned. Score:  66
Sequences (2:5) Aligned. Score:  71
Sequences (2:6) Aligned. Score:  42
Sequences (2:7) Aligned. Score:  100
Sequences (2:8) Aligned. Score:  71
Sequences (2:9) Aligned. Score:  67
Sequences (2:10) Aligned. Score:  65
Sequences (2:11) Aligned. Score:  71
Sequences (3:4) Aligned. Score:  94
Sequences (3:5) Aligned. Score:  66
Sequences (3:6) Aligned. Score:  42
Sequences (3:7) Aligned. Score:  65
Sequences (3:8) Aligned. Score:  66
Sequences (3:9) Aligned. Score:  66
Sequences (3:10) Aligned. Score:  94
Sequences (3:11) Aligned. Score:  66
Sequences (4:5) Aligned. Score:  66
Sequences (4:6) Aligned. Score:  41
Sequences (4:7) Aligned. Score:  66
Sequences (4:8) Aligned. Score:  65
Sequences (4:9) Aligned. Score:  66
Sequences (4:10) Aligned. Score:  93
Sequences (4:11) Aligned. Score:  65
Sequences (5:6) Aligned. Score:  41
Sequences (5:7) Aligned. Score:  71
Sequences (5:8) Aligned. Score:  93
Sequences (5:9) Aligned. Score:  66
Sequences (5:10) Aligned. Score:  66
Sequences (5:11) Aligned. Score:  93
Sequences (6:7) Aligned. Score:  42
Sequences (6:8) Aligned. Score:  42
Sequences (6:9) Aligned. Score:  39
Sequences (6:10) Aligned. Score:  41
Sequences (6:11) Aligned. Score:  42
Sequences (7:8) Aligned. Score:  71
Sequences (7:9) Aligned. Score:  67
Sequences (7:10) Aligned. Score:  65
Sequences (7:11) Aligned. Score:  71
Sequences (8:9) Aligned. Score:  66
Sequences (8:10) Aligned. Score:  66
Sequences (8:11) Aligned. Score:  100
Sequences (9:10) Aligned. Score:  66
Sequences (9:11) Aligned. Score:  66
Sequences (10:11) Aligned. Score:  66
Guide tree        file created:   [baaBsaWDA.dnd]
Start of Multiple Alignment
There are 10 groups
Aligning...
Group 1: Sequences:   2      Score:8554
Group 2: Sequences:   3      Score:8520
Group 3: Sequences:   4      Score:7773
Group 4: Sequences:   2      Score:8823
Group 5: Sequences:   2      Score:8860
Group 6: Sequences:   3      Score:8533
Group 7: Sequences:   5      Score:8317
Group 8: Sequences:   9      Score:7685
Group 9: Sequences:   2      Score:8364
Group 10: Sequences:  11      Score:3915
Alignment Score 113614
CLUSTAL-Alignment file created  [baaBsaWDA.aln]
CLUSTAL W (1.7) multiple sequence alignment


g31377618       --MARTGWTSPIPLCVSLLLTCG-FAEAGKLLVVPMDGSHWFTMQSVVEKLILRGHEVVV
g29789078       --MARAGWTSPVPLCVCLLLTCG-FAEAGKLLVVPMDGSHWFTMQSVVEKLILRGHEVVV
g41282213       --MARAGWTGLLPLYVCLLLTCG-FAKAGKLLVVPMDGSHWFTMQSVVEKLILRGHEVVV
g4507815        -MACLLRSFQRISAGVFFLALWG-MVVGDKLLVVPQDGSHWLSMKDIVEVLSDRGHEIVV
```

```
g136729         MAVESQGGRP-LVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSMLGAIQQLQQRGHEIVV
g8850236        MAVESQGGRP-LVLGLLLCVLGPVVSHAGKILLIPVDGSHWLSMLGAIQQLQQRGHEIVV
g6005930        MARGLQVPLPRLATGLLLLLSVQPWAESGKVLVVPTDGSPWLSMREALRELHARGHQAVV
g136731         MARGLQVPLPRLATGLLLLLSVQPWAESGKVLVVPTDGSPWLSMREALRELHARGHQAVV
g13487900       MATGLQVPLPWLATGLLLLLSVQPWAESGKVLVVPIDGSHWLSMREVLRELHARGHQAVV
2912330CD1      --MSMKWTSALLLIQLSCYFSSG--S-CGKVLVWPTEFSHWMNIKTILDELVQRGHEVTV
g10863941       --MSMKWTSALLLIQLSCYFSSG--S-CGKVLVWPTEFSHWMNIKTILDELVQRGHEVTV
                              :   :             .*:*: * : * *:.:   :  *  ***: .*

g31377618       VMPEVSWQLGKSL--NC TVKTYSTSYTLEDLDREFMDFADAQWKA--QVRSLFSLFLSSS
g29789078       VMPEVSWQLERSL--NCTVKTYSTSYTLEDQNREFMVFAHAQWKA--QAQSIFSLLMSSS
g41282213       VMPEVSWQLGRSL--NCTVKTYSTSYTLEDQDREFMVFADARWTA--PLRSAFSLLTSSS
g4507815        VVPEVNLLLKEYK--YYTRKIYPVPYDQEELKNRYQSFGNNHFAE--RSFLTAPQTEYRN
g136729         LAPDASLYIRDGA--FYTLKTYPVPFQREDVKESFVSLGHNVFEN--DSFLQRVIKTYKK
g8850236        LAPDASLYIRDGA--FYTLKTYPVPFQREDVKESFVSLGHNVFEN--DSFLQRVIKTYKK
g6005930        LTPEVNMHIKEEK--FFTLTAYAVPWTQKEFDRVTLGYTQGFFET--EHLLKRYSRSMAI
g136731         LTPEVNMHIKEEK--FFTLTAYAVPWTQKEFDRVTLGYTQGFFET--EHLLKRYSRSMAI
g13487900       LTPEVNMHIKEEN--FFTLTTYAISWTQDEFDRHVLGHTQLYFET--EHFLKKFFRSMAM
2912330CD1      LASSASISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPKDTFWSYFSQVQEI
g10863941       LASSASISFDPNSPSTLKFEVYPVSLTKTEFEDIIKQLVKRWAELPKDTFWSYFSQVQEI
                : ....  :         .   *. .    : .        .

g31377618       NG-FFNLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDACGLIVAKYFSLPSVVFARGI
g29789078       SG-FLDLFFSHCRSLFNDRKLVEYLKESSFDAVFLDPFDTCGLIVAKYFSLPSVVFTRGI
g41282213       NG-IFDLFFSNCRSLFNDRKLVEYLKESCFDAVFLDPFDACGLIVAKYFSLPSVVFARGI
g4507815        NMIVIGLYFINCQSLLQDRDTLNFFKESKFDALFTDPALPCGVILAEYLGLPSVYLFRGF
g136729         IKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIVAQYLSLPTVFFLHAL
g8850236        IKKDSAMLLSGCSHLLHNKELMASLAESSFDVMLTDPFLPCSPIVAQYLSLPTVFFLHAL
g6005930        MNNVSLALHRCCVELLHNEALIRHLNATSFDVVLTDPVNLCGAVLAKYLSIPAVFFWRYI
g136731         MNNVSLALHRCCVELLHNEALIRHLNATSFDVVLTDPVNLCGAVLAKYLSIPAVFFWRYI
g13487900       LNNMSLVYHRSCVELLHNEALIRHLNATSFDVVLTDPVNLCAAVLAKYLSIPTVFFLRNI
2912330CD1      MWTFNDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYSLRFS
g10863941       MWTFNDILRKFCKDIVSNKKLMKKLQESRFDVVLADAVFPFGELLAELLKIPFVYRPRFS
                        *  :. :.  :  :  : **.:: *.    . ::*: : :* *   :

g31377618       ACHYLEEGAQ-CPAPLSYVPRILLGFSDAMTFKERVRNHIMHLEEHLFCQYF-SKNALEI
g29789078       FCHHLEEGAQ-CPAPLSYVPNDLLGFSDAMTFKERVWNHIVHLEDHLFCQYL-FRNALEI
g41282213       FCHYLEEGAQ-CPAPLSYVPRLLLGFSDAMTFKERVWNHIMHLEEHLFCPYF-FKNVLEI
g4507815        PCSLEHTFSR-SPDPVSYIPRCYTKFSDHMTFSQRVANFLVNLLEPYLFYCL-FSKYEKL
g136729         PCSLEFEATQ-CPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFSQNFLCDVV-YSPYATL
g8850236        PCSLEFEATQ-CPNPFSYVPRPLSSHSDHMTFLQRVKNMLIAFSQNFLCDVV-YSPYATL
g6005930        PCDLDFKGTQ-CPNPSSYIPKLLTTNSDHMTFLQRVKNMLYPLALSYICHTF-SAPYASL
g136731         PCDLDFKGTQ-CPNPSSYIPKLLTTNSDHMTFLQRVKNMLYPLALSYICHTF-SAPYASL
g13487900       PCDLDFKGTQ-CPNPSSYIPRLLTTNSDHMTFMQRVKNMLYPLALSYICHAF-SAPYASL
2912330CD1      PGYAIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQF
g10863941       PGYAIEKHSGGLLFPPSYVPVVMSELSDQMTFIERVKNMIYVLYFEFWFQIFDMKKWDQF
                         :     * **:*       ** *** :** * :  :         .       :

g31377618       ASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGINCHQGKPLPMEFEAY
g29789078       ASEILQTPVTAYDLYSHTSIWLLRTDFVLDYPKPVMPNMIFIGGINCHQGKPLPMEFEAY
g41282213       ASEILQTPVTAYDLYSHTSIWLLRTDFVLEYPKPVMPNMIFIGGINCHQGKPVPMEFEAY
g4507815        ASAVLKRDVDIITL-SEVSVWLLRYDFVLEYPRPVMPNMVFIGGINCKKRKDLSQEFEAY
g136729         ASEFLQREVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVGGINCLHQNPLSQEFEAY
g8850236        ASEFLQREVTVQDLLSSASVWLFRSDFVKDYPRPIMPNMVFVGGINCLHQNPLSQEFEAY
g6005930        ASELFQREVSVVDLVSYASVWLFRGDFVMDYPRPIMPNMVFIGGINCANGKPLSQEFEAY
g136731         ASELFQREVSVVDLVSYASVWLFRGDFVMDYPRPIMPNMVFIGGINCANGKPLSQEFEAY
g13487900       ASELFQREVSVVDILSHASVWLFRGDFVMDYPRPIMPNMVFIGGINCANRKPLSQEFEAY
2912330CD1      YSEVLGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEF
g10863941       YSEVLGRPTTLSETMAKADIWLIRNYWDFQFPHPLLPNVEFVGGLHCKPAKPLPKEMEEF
                 * .:   .       : ..:**:*  :  ::*:*::**: *:**::*   : :. *:* :

g31377618       INASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVK
g29789078       INASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVK
g41282213       INASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVK
g4507815        INASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKNPQTVLWRYTGTRPSNLANNTILVK
g136729         INASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVK
g8850236        INASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVK
```

```
g6005930        INASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVK
g136731         INASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVK
g13487900       INASGEHGIVVFSLGSMVSEIPEKKAMAIADALGKIPQTVLWRYTGTRPSNLANNTILVK
2912330CD1      VQSSGENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYK
g10863941       VQSSGENGVVVFSLGSMVSNTSEERANVIASALAKIPQKVLWRFDGNKPDTLGLNTRLYK
                :::***:*:**********: .*::* .**.**.* **.****: *.:*..*. ** * *

g31377618       WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTL
g29789078       WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTL
g41282213       WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTL
g4507815        WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTL
g136729         WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTL
g8850236        WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTL
g6005930        WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTL
g136731         WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTL
g13487900       WLPQNDLLGHPMTRAFITHAGSHGVYESICNGVPMVMMPLFGDQMDNAKRMETKGAGVTL
2912330CD1      WIPQNDLLGHPKTKAFITHGGMNGIYEAIYHGVPMVGVPIFGDQLDNIAHMKAKGAAVEI
g10863941       WIPQNDLLGHPKTRAFITHGGANGIYKAISPRIPMVGVPLFADQPDNIAHMKAKGAAVSL
                *:********** *:*****.* :*:*::*   :*** :*:*.** **  :*::***.* :

g31377618       NVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPH
g29789078       NVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPH
g41282213       NVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPH
g4507815        NVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPH
g136729         NVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPH
g8850236        NVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPH
g6005930        NVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPH
g136731         NVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPH
g13487900       NVLEMTSEDLENALKAVINDKSYKENIMRLSSLHKDRPVEPLDLAVFWVEFVMRHKGAPH
2912330CD1      NFKTMTSEDLLRALRTVITDSSYKENAMRLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKH
g10863941       DFHTMSSTDLLNALKTVINDPLYKENAMKLSRIHHDQPVKPLDRAVFWIEFVMRHKGAKH
                :.  *:* ** .**::**.*  **** *:** :*:*:**:*** ****:********* *

g31377618       LRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
g29789078       LRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
g41282213       LRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
g4507815        LRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCPYGYPKCLGKKGRVKKAHKSKTH
g136729         LRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
g8850236        LRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
g6005930        LRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
g136731         LRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
g13487900       LRPAAHDLTWYQYHSLDVIGFLLAVVLTVAFITFKCCAYGYRKCLGKKGRVKKAHKSKTH
2912330CD1      LRSAAHDLTWFQHYSIDVIGFLLTCVATAIFLFTKCFLFSCQKFNKTR-KIEKRE-----
g10863941       LRVAAHDLTWFQYHSLDVTGFLLACVATVIFIITKC-LFCVWKFVRTG-KKGKRD-----
                ** *******:*::*:** ****: * *. *:  **  :   *   .  :  * .
```

Submit sequences to: BLAST2 Submit Reset

IncyteGenomics

1: Pharmacogenetics. 1997 Aug;7(4):255-69.

The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence.

Mackenzie PI, Owens IS, Burchell B, Bock KW, Bairoch A, Belanger A, Fournel-Gigleux S, Green M, Hum DW, Iyanagi T, Lancet D, Louisot P, Magdalou J, Chowdhury JR, Ritter JK, Schachter H, Tephly TR, Tipton KF, Nebert DW.

Department of Clinical Pharmacology, Flinders University of South Australia, Bedford Park.

This review represents an update of the nomenclature system for the UDP glucuronosyltransferase gene superfamily, which is based on divergent evolution. Since the previous review in 1991, sequences of many related UDP glycosyltransferases from lower organisms have appeared in the database, which expand our database considerably. At latest count, in animals, yeast, plants and bacteria there are 110 distinct cDNAs/genes whose protein products all contain a characteristic 'signature sequence' and, thus, are regarded as members of the same superfamily. Comparison of a relatedness tree of proteins leads to the definition of 33 families. It should be emphasized that at least six cloned UDP-GlcNAc N-acetylglucosaminyltransferases are not sufficiently homologous to be included as members of this superfamily and may represent an example of convergent evolution. For naming each gene, it is recommended that the root symbol UGT for human (Ugt for mouse and Drosophila), denoting 'UDP glycosyltransferase,' be followed by an Arabic number representing the family, a letter designating the subfamily, and an Arabic numeral denoting the individual gene within the family or subfamily, e.g. 'human UGT2B4' and 'mouse Ugt2b5'. We recommend the name 'UDP glycosyltransferase' because many of the proteins do not preferentially use UDP glucuronic acid, or their nucleotide sugar preference is unknown. Whereas the gene is italicized, the corresponding cDNA, transcript, protein and enzyme activity should be written with upper-case letters and without italics, e.g. 'human or mouse UGT1A1.' The UGT1 gene (spanning > 500 kb) contains at least 12 promoters/first exons, which can be spliced and joined with common exons 2 through 5, leading to different N-terminal halves but identical C-terminal halves of the gene products; in this scheme each first exon is regarded as a distinct gene (e.g. UGT1A1, UGT1A2, ... UGT1A12). When an orthologous gene between species cannot be identified with certainty, as occurs in the UGT2B subfamily, sequential naming of the genes is being carried out chronologically as they become characterized. We suggest that the Human Gene Nomenclature Guidelines

(http://www.gene.acl.ac.uk/nomenclature/guidelines.html++ +) be used for all species other than the mouse and Drosophila. Thirty published human UGT1A1 mutant alleles responsible for clinical hyperbilirubinemias are listed herein, and given numbers following an asterisk (e.g. UGT1A1*30) consistent with the Human Gene Nomenclature Guidelines. It is anticipated that this UGT gene nomenclature system will require updating on a regular basis.

Publication Types:
- Review
- Review, Tutorial

MeSH Terms:
- Amino Acid Sequence
- Animals
- Evolution, Molecular*
- Genes, Structural*
- Glucuronosyltransferase/chemistry
- Glucuronosyltransferase/genetics*
- Human
- Molecular Sequence Data
- Multigene Family*
- Sequence Alignment
- Sequence Homology, Amino Acid
- Support, U.S. Gov't, P.H.S.
- Terminology*

Substances:
- Glucuronosyltransferase

Grant Support:
- P30 ES06096/ES/NIEHS

PMID: 9295054 [PubMed - indexed for MEDLINE]

1: P22309. UDP-glucuronosylt...[gi:136729]


```
LOCUS       P22309                  533 aa            linear   PRI 15-MAR-2004
DEFINITION  UDP-glucuronosyltransferase 1-1 precursor, microsomal
            (UDP-glucuronosyltransferase 1A1) (UDPGT) (UGT1*1) (UGT1-01)
            (UGT1.1) (UGT-1A) (UGT1A) (Bilirubin specific UDPGT isozyme 1)
            (HUG-BR1).
ACCESSION   P22309
VERSION     P22309  GI:136729
DBSOURCE    swissprot: locus UD11_HUMAN, accession P22309;
            class: standard.
            created: Aug 1, 1991.
            sequence updated: Aug 1, 1991.
            annotation updated: Mar 15, 2004.
            xrefs: gi: 340131, gi: 340132, gi: 340129, gi: 459838, gi: 340127,
            gi: 340128, gi: 184472, gi: 184473, gi: 11118740, gi: 11118749, gi:
            5732165, gi: 6094671, gi: 3059176, gi: 3059177, gi: 87534
            xrefs (non-sequence databases): GenewHGNC:12530, MIM 191740, MIM
            143500, MIM 218800, MIM 606785, GO0006789, GO0008210,
            InterProIPR002213, PfamPF00201, PROSITEPS00375
KEYWORDS    Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
            Signal; Multigene family; Microsome; Alternative splicing; Disease
            mutation.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 533)
  AUTHORS   Ritter,J.K., Crawford,J.M. and Owens,I.S.
  TITLE     Cloning of two human liver bilirubin UDP-glucuronosyltransferase
            cDNAs with expression in COS-1 cells
  JOURNAL   J. Biol. Chem. 266 (2), 1043-1047 (1991)
  MEDLINE   91093210
   PUBMED   1898728
  REMARK    SEQUENCE FROM N.A.
            TISSUE=Liver
REFERENCE   2  (residues 1 to 533)
  AUTHORS   Ritter,J.K., Chen,F., Sheen,Y.Y., Tran,H.M., Kimura,S.,
            Yeatman,M.T. and Owens,I.S.
  TITLE     A novel complex locus UGT1 encodes human bilirubin, phenol, and
            other UDP-glucuronosyltransferase isozymes with identical carboxyl
            termini
  JOURNAL   J. Biol. Chem. 267 (5), 3257-3261 (1992)
  MEDLINE   92147680
   PUBMED   1339448
  REMARK    SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
REFERENCE   3  (residues 1 to 533)
  AUTHORS   Gong,Q.H., Cho,J.W., Huang,T., Potter,C., Gholami,N., Basu,N.K.,
            Kubota,S., Carvalho,S., Pennington,M.W., Owens,I.S. and
            Popescu,N.C.
  TITLE     Thirteen UDPglucuronosyltransferase genes are encoded at the human
            UGT1 gene complex locus
  JOURNAL   Pharmacogenetics 11 (4), 357-368 (2001)
  MEDLINE   21327373
   PUBMED   11434514
  REMARK    SEQUENCE FROM N.A.
REFERENCE   4  (residues 1 to 533)
```

AUTHORS Gattung,S., Stoneking,T. and Davidson,T.
TITLE Direct Submission
JOURNAL Submitted (~MAR-1999)
REMARK SEQUENCE FROM N.A.
REFERENCE 5 (residues 1 to 533)
AUTHORS Ueyama,H., Koiwai,O., Soeda,Y., Sato,H., Satoh,Y., Ohkubo,I. and Doida,Y.
TITLE Analysis of the promoter of human bilirubin UDP-glucuronosyltransferase gene (UGT1*1) in relevance to Gilbert's syndrome
JOURNAL Hepatol. Res. 9, 152-163 (1997)
REMARK SEQUENCE OF 1-50 FROM N.A.
REFERENCE 6 (residues 1 to 533)
AUTHORS Bosma,P.J., Chowdhury,J.R., Huang,T.J., Lahiri,P., Elferink,R.P., Van Es,H.H., Lederstein,M., Whitington,P.F., Jansen,P.L. and Chowdhury,N.R.
TITLE Mechanisms of inherited deficiencies of multiple UDP-glucuronosyltransferase isoforms in two patients with Crigler-Najjar syndrome, type I
JOURNAL FASEB J. 6 (10), 2859-2863 (1992)
MEDLINE 92339803
PUBMED 1634050
REMARK VARIANT CN-I PHE-375.
REFERENCE 7 (residues 1 to 533)
AUTHORS Aono,S., Yamada,Y., Keino,H., Hanada,N., Nakagawa,T., Sasaoka,Y., Yazawa,T., Sato,H. and Koiwai,O.
TITLE Identification of defect in the genes for bilirubin UDP-glucuronosyl-transferase in a patient with Crigler-Najjar syndrome type II
JOURNAL Biochem. Biophys. Res. Commun. 197 (3), 1239-1244 (1993)
MEDLINE 94107323
PUBMED 8280139
REMARK VARIANTS CN-II ARG-71 AND ASP-486.
REFERENCE 8 (residues 1 to 533)
AUTHORS Moghrabi,N., Clarke,D.J., Boxer,M. and Burchell,B.
TITLE Identification of an A-to-G missense mutation in exon 2 of the UGT1 gene complex that causes Crigler-Najjar syndrome type 2
JOURNAL Genomics 18 (1), 171-173 (1993)
MEDLINE 94102756
PUBMED 8276413
REMARK VARIANT CN-II ARG-331.
REFERENCE 9 (residues 1 to 533)
AUTHORS Ritter,J.K., Yeatman,M.T., Kaiser,C., Gridelli,B. and Owens,I.S.
TITLE A phenylalanine codon deletion at the UGT1 gene complex locus of a Crigler-Najjar type I patient generates a pH-sensitive bilirubin UDP-glucuronosyltransferase
JOURNAL J. Biol. Chem. 268 (31), 23573-23579 (1993)
MEDLINE 94043159
PUBMED 8226884
REMARK VARIANT CN-I PHE-170 DEL.
REFERENCE 10 (residues 1 to 533)
AUTHORS Labrune,P., Myara,A., Hadchouel,M., Ronchi,F., Bernard,O., Trivin,F., Chowdhury,N.R., Chowdhury,J.R., Munnich,A. and Odievre,M.
TITLE Genetic heterogeneity of Crigler-Najjar syndrome type I: a study of 14 cases
JOURNAL Hum. Genet. 94 (6), 693-697 (1994)
MEDLINE 95080780
PUBMED 7989045
REMARK VARIANTS CN-I VAL-292; GLU-308; ARG-357; THR-368; ARG-381; PRO-401 AND GLU-428.
REFERENCE 11 (residues 1 to 533)
AUTHORS Seppen,J., Bosma,P.J., Goldhoorn,B.G., Bakker,C.T., Chowdhury,J.R., Chowdhury,N.R., Jansen,P.L. and Oude Elferink,R.P.
TITLE Discrimination between Crigler-Najjar type I and II by expression of mutant bilirubin uridine diphosphate-glucuronosyltransferase
JOURNAL J. Clin. Invest. 94 (6), 2385-2391 (1994)
MEDLINE 95081424
PUBMED 7989595

```
  REMARK    VARIANTS CN GLU-175; ARG-177; TRP-209; ARG-276 AND PHE-375.
REFERENCE   12 (residues 1 to 533)
  AUTHORS   Aono,S., Adachi,Y., Uyama,E., Yamada,Y., Keino,H., Nanno,T.,
            Koiwai,O. and Sato,H.
  TITLE     Analysis of genes for bilirubin UDP-glucuronosyltransferase in
            Gilbert's syndrome
  JOURNAL   Lancet 345 (8955), 958-959 (1995)
  MEDLINE   95231122
   PUBMED   7715297
  REMARK    VARIANTS GILBERT SYNDROME ARG-71; GLN-229 AND GLY-367.
            TISSUE=Liver, and Peripheral blood leukocytes
REFERENCE   13 (residues 1 to 533)
  AUTHORS   Yamamoto,K., Soeda,Y., Kamisako,T., Hosaka,H., Fukano,M., Sato,H.,
            Fujiyama,Y., Adachi,Y., Satoh,Y. and Bamba,T.
  TITLE     Analysis of bilirubin uridine 5'-diphosphate
            (UDP)-glucuronosyltransferase gene mutations in seven patients with
            Crigler-Najjar syndrome type II
  JOURNAL   J. Hum. Genet. 43 (2), 111-114 (1998)
  MEDLINE   98284535
   PUBMED   9621515
  REMARK    VARIANTS CN-II ARG-71; TRP-209; GLN-229 AND ASP-486.
REFERENCE   14 (residues 1 to 533)
  AUTHORS   Maruo,Y., Sato,H., Yamano,T., Doida,Y. and Shimada,M.
  TITLE     Gilbert syndrome caused by a homozygous missense mutation
            (Tyr486Asp) of bilirubin UDP-glucuronosyltransferase gene
  JOURNAL   J. Pediatr. 132 (6), 1045-1047 (1998)
  MEDLINE   98291073
   PUBMED   9627603
  REMARK    VARIANT GILBERT SYNDROME ASP-486.
COMMENT     ---------------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------------.
            [FUNCTION] UDPGT is of major importance in the conjugation and
            subsequent elimination of potentially toxic xenobiotics and
            endogenous compounds. This isoform glucuronidates bilirubin
            IX-alpha to form both the IX-alpha-C8 and IX-alpha-C12
            monoconjugates and diconjugate.
            [CATALYTIC ACTIVITY] UDP-glucuronate + acceptor = UDP + acceptor
            beta-D-glucuronoside.
            [SUBCELLULAR LOCATION] Microsomal.
            [ALTERNATIVE PRODUCTS] Event=Alternative splicing; Named
            isoforms=1; Comment=A number of isoforms are produced. The
            different isozymes have a different N-terminal domain and a common
            C-terminal domain of 245 residues; Name=1; IsoId=P22309-1;
            Sequence=Displayed.
            [TISSUE SPECIFICITY] Expressed in liver. Not expressed in skin or
            kidney.
            [DISEASE] Defects in UGT1A1 are the cause of Gilbert syndrome
            [MIM:143500]. Gilbert syndrome occurs as a consequence of reduced
            bilirubin transferase activity and is often detected in young
            adults with vague nonspecific complaints.
            [DISEASE] Defects in UGT1A1 are the cause of Crigler-Najjar
            syndrome type I (CN-I) [MIM:218800]. CN-I patients have severe
            hyperbilirubinemia and usually die of kernicterus (bilirubin
            accumulation in the basal ganglia and brainstem nuclei) within the
            first year of life. CN-I inheritance is autosomal recessive.
            [DISEASE] Defects in UGT1A1 are the cause of Crigler-Najjar
            syndrome type II (CN-II) [MIM:606785]. CN-II patients have less
            severe hyperbilirubinemia and usually survive into adulthood
            without neurologic damage. Phenobarbital, which induces the
            partially deficient glucuronyl transferase, can diminish the
            jaundice. CN-II inheritance is autosomal dominant.
            [SIMILARITY] Belongs to the UDP-glycosyltransferase family.
FEATURES             Location/Qualifiers
     source          1..533
```

```
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
gene            1..533
                /gene="UGT1A1"
                /note="synonyms: UGT1, GNT1"
Protein         1..533
                /gene="UGT1A1"
                /product="UDP-glucuronosyltransferase 1-1 precursor,
                microsomal"
                /EC_number="2.4.1.17"
Region          1..25
                /gene="UGT1A1"
                /region_name="Signal"
                /note="Potential."
Region          26..533
                /gene="UGT1A1"
                /region_name="Mature chain"
                /note="UDP-GLUCURONOSYLTRANSFERASE 1-1."
Region          71
                /gene="UGT1A1"
                /region_name="Variant"
                /note="G -> R (in CN-II and Gilbert syndrome).
                /FTId=VAR_009504."
Site            102
                /gene="UGT1A1"
                /site_type="glycosylation"
                /note="N-linked (GlcNAc...) (Potential)."
Region          170
                /gene="UGT1A1"
                /region_name="Variant"
                /note="Missing (in CN-I; has nearly normal activity at pH
                7.6 and is inactive at pH 6.4). /FTId=VAR_007695."
Region          175
                /gene="UGT1A1"
                /region_name="Variant"
                /note="L -> E (in CN-II). /FTId=VAR_007696."
Region          177
                /gene="UGT1A1"
                /region_name="Variant"
                /note="C -> R (in CN-I). /FTId=VAR_007697."
Region          209
                /gene="UGT1A1"
                /region_name="Variant"
                /note="R -> W (in CN-II). /FTId=VAR_007698."
Region          229
                /gene="UGT1A1"
                /region_name="Variant"
                /note="P -> Q (in CN-II and Gilbert syndrome).
                /FTId=VAR_009505."
Region          276
                /gene="UGT1A1"
                /region_name="Variant"
                /note="G -> R (in CN-I). /FTId=VAR_007699."
Region          292
                /gene="UGT1A1"
                /region_name="Variant"
                /note="A -> V (in CN-I). /FTId=VAR_007700."
Site            295
                /gene="UGT1A1"
                /site_type="glycosylation"
                /note="N-linked (GlcNAc...) (Potential)."
Region          308
                /gene="UGT1A1"
                /region_name="Variant"
                /note="G -> E (in CN-I). /FTId=VAR_007701."
Region          331
                /gene="UGT1A1"
                /region_name="Variant"
                /note="Q -> R (in CN-II). /FTId=VAR_007702."
```

```
     Site            347
                     /gene="UGT1A1"
                     /site_type="glycosylation"
                     /note="N-linked (GlcNAc...) (Potential)."
     Region          357
                     /gene="UGT1A1"
                     /region_name="Variant"
                     /note="Q -> R (in CN-I). /FTId=VAR_007703."
     Region          367
                     /gene="UGT1A1"
                     /region_name="Variant"
                     /note="R -> G (in Gilbert syndrome). /FTId=VAR_012283."
     Region          368
                     /gene="UGT1A1"
                     /region_name="Variant"
                     /note="A -> T (in CN-I). /FTId=VAR_007704."
     Region          375
                     /gene="UGT1A1"
                     /region_name="Variant"
                     /note="S -> F (in CN-I). /FTId=VAR_007705."
     Region          381
                     /gene="UGT1A1"
                     /region_name="Variant"
                     /note="S -> R (in CN-I). /FTId=VAR_007706."
     Region          401
                     /gene="UGT1A1"
                     /region_name="Variant"
                     /note="A -> P (in CN-I). /FTId=VAR_007707."
     Region          428
                     /gene="UGT1A1"
                     /region_name="Variant"
                     /note="K -> E (in CN-I). /FTId=VAR_007708."
     Region          486
                     /gene="UGT1A1"
                     /region_name="Variant"
                     /note="Y -> D (in CN-II and Gilbert syndrome).
                     /FTId=VAR_007709."
     Region          491..507
                     /gene="UGT1A1"
                     /region_name="Transmembrane region"
                     /note="Potential."
ORIGIN
        1 mavesqggrp lvlglllcvl gpvvshagki llipvdgshw lsmlgaiqql qqrgheivvl
       61 apdaslyird gafytlktyp vpfqredvke sfvslghnvf endsflqrvi ktykkikkds
      121 amllsgcshl lhnkelmasl aessfdvmlt dpflpcspiv aqylslptvf flhalpcsle
      181 featqcpnpf syvprplssh sdhmtflqrv knmliafsqn flcdvvyspy atlaseflqr
      241 evtvqdllss asvwlfrsdf vkdyprpimp nmvfvgginc lhqnplsqef eayinasgeh
      301 givvfslgsm vseipekkam aiadalgkip qtvlwrytgt rpsnlannti lvkwlpqndl
      361 lghpmtrafi thagshgvye sicngvpmvm mplfgdqmdn akrmetkgag vtlnvlemts
      421 edlenalkav indksykeni mrlsslhkdr pvepldlavf wvefvmrhkg aphlrpaahd
      481 ltwyqyhsld vigfllavvl tvafitfkcc aygyrkclgk kgrvkkahks kth
//
```

2: NP_061949. UDP glycosyltrans...[gi:31377618]

BLink, Domains, Links

```
LOCUS       NP_061949                 530 aa            linear   PRI 22-DEC-2003
DEFINITION  UDP glycosyltransferase 1 family, polypeptide A8 [Homo sapiens].
ACCESSION   NP_061949
VERSION     NP_061949.3  GI:31377618
DBSOURCE    REFSEQ: accession NM_019076.3
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 530)
  AUTHORS   Gregory,P.A., Gardner-Stephen,D.A., Lewinsky,R.H., Duncliffe,K.N.
            and Mackenzie,P.I.
  TITLE     Cloning and characterization of the human
```

```
            UDP-glucuronosyltransferase 1A8, 1A9, and 1A10 gene promoters:
            differential regulation through an interior-like region
  JOURNAL   J. Biol. Chem. 278 (38), 36107-36114 (2003)
   PUBMED   12847094
  REMARK    GeneRIF: UGT1A8, 1A9, and 1A10 genes are differentially regulated
            through an initiator element in their 5'-flanking regions
REFERENCE   2  (residues 1 to 530)
  AUTHORS   Huang,Y.H., Galijatovic,A., Nguyen,N., Geske,D., Beaton,D.,
            Green,J., Green,M., Peters,W.H. and Tukey,R.H.
  TITLE     Identification and functional characterization of
            UDP-glucuronosyltransferases UGT1A8*1, UGT1A8*2 and UGT1A8*3
  JOURNAL   Pharmacogenetics 12 (4), 287-297 (2002)
   PUBMED   12042666
REFERENCE   3  (residues 1 to 530)
  AUTHORS   Gong,Q.H., Cho,J.W., Huang,T., Potter,C., Gholami,N., Basu,N.K.,
            Kubota,S., Carvalho,S., Pennington,M.W., Owens,I.S. and
            Popescu,N.C.
  TITLE     Thirteen UDPglucuronosyltransferase genes are encoded at the human
            UGT1 gene complex locus
  JOURNAL   Pharmacogenetics 11 (4), 357-368 (2001)
   PUBMED   11434514
REFERENCE   4  (residues 1 to 530)
  AUTHORS   Strassburg,C.P., Manns,M.P. and Tukey,R.H.
  TITLE     Expression of the UDP-glucuronosyltransferase 1A locus in human
            colon. Identification and characterization of the novel
            extrahepatic UGT1A8
  JOURNAL   J. Biol. Chem. 273 (15), 8719-8726 (1998)
   PUBMED   9535849
REFERENCE   5  (residues 1 to 530)
  AUTHORS   Mackenzie,P.I., Owens,I.S., Burchell,B., Bock,K.W., Bairoch,A.,
            Belanger,A., Fournel-Gigleux,S., Green,M., Hum,D.W., Iyanagi,T.,
            Lancet,D., Louisot,P., Magdalou,J., Chowdhury,J.R., Ritter,J.K.,
            Schachter,H., Tephly,T.R., Tipton,K.F. and Nebert,D.W.
  TITLE     The UDP glycosyltransferase gene superfamily: recommended
            nomenclature update based on evolutionary divergence
  JOURNAL   Pharmacogenetics 7 (4), 255-269 (1997)
   PUBMED   9295054
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from AF462268.1.
            On Jun 4, 2003 this sequence version replaced gi:19424142.
FEATURES             Location/Qualifiers
     source          1..530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
     Protein         1..530
                     /product="UDP glycosyltransferase 1 family, polypeptide
                     A8"
     variation       15
                     /replace="C"
                     /replace="S"
                     /db_xref="dbSNP:1126783"
     Region          26..522
                     /region_name="UDP-glucoronosyl and UDP-glucosyl
                     transferase"
                     /note="UDPGT"
                     /db_xref="CDD:22944"
     variation       95
                     /replace="H"
                     /replace="D"
                     /db_xref="dbSNP:1126785"
     variation       105
                     /replace="L"
                     /replace="M"
                     /db_xref="dbSNP:1126788"
     variation       109
                     /replace="L"
                     /replace="F"
```

```
                     /db_xref="dbSNP:1126792"
     variation       110
                     /replace="L"
                     /replace="M"
                     /db_xref="dbSNP:1126793"
     variation       118
                     /replace="D"
                     /replace="N"
                     /db_xref="dbSNP:1126798"
     variation       208
                     /replace="R"
                     /replace="W"
                     /db_xref="dbSNP:1126802"
     variation       212
                     /replace="M"
                     /replace="V"
                     /db_xref="dbSNP:1126803"
     variation       216
                     /replace="D"
                     /replace="E"
                     /db_xref="dbSNP:1126804"
     variation       508
                     /replace="P"
                     /replace="A"
                     /db_xref="dbSNP:1042709"
     CDS             1..530
                     /gene="UGT1A8"
                     /coded_by="NM_019076.3:64..1656"
                     /note="go_function: UDP-glycosyltransferase activity [goid
                     0008194] [evidence P] [pmid 9295054];
                     go_process: metabolism [goid 0008152] [evidence P] [pmid
                     9295054]"
                     /db_xref="GeneID:54576"
                     /db_xref="LocusID:54576"
                     /db_xref="MIM:606433"
ORIGIN
        1 martgwtspi plcvsllltc gfaeagkllv vpmdgshwft mqsvveklil rghevvvvmp
       61 evswqlgksl nctvktysts ytledldref mdfadaqwka qvrslfslfl sssngffnlf
      121 fshcrslfnd rklveylkes sfdavfldpf dacglivaky fslpsvvfar giachyleeg
      181 aqcpaplsyv prillgfsda mtfkervrnh imhleehlfc qyfsknalei aseilqtpvt
      241 aydlyshtsi wllrtdfvld ypkpvmpnmi figginchqg kplpmefeay inasgehgiv
      301 vfslgsmvse ipekkamaia dalgkipqtv lwrytgtrps nlanntilvk wlpqndllgh
      361 pmtrafitha gshgvyesic ngvpmvmmpl fgdqmdnakr metkgagvtl nvlemtsedl
      421 enalkavind ksykenimrl sslhkdrpve pldlavfwve fvmrhkgaph lrpaahdltw
      481 yqyhsldvig fllavvltva fitfkccayg yrkclgkkgr vkkahkskth
//
```

3: NP_061948. UDP glycosyltrans...[gi:29789078] BLink, Domains, Links

```
LOCUS       NP_061948                530 aa            linear   PRI 21-DEC-2003
DEFINITION  UDP glycosyltransferase 1 family, polypeptide A10;
            UDP-glucuronosyltransferase 1A10 [Homo sapiens].
ACCESSION   NP_061948
VERSION     NP_061948.1  GI:29789078
DBSOURCE    REFSEQ: accession NM_019075.1
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 530)
  AUTHORS   Gregory,P.A., Gardner-Stephen,D.A., Lewinsky,R.H., Duncliffe,K.N.
            and Mackenzie,P.I.
  TITLE     Cloning and characterization of the human
            UDP-glucuronosyltransferase 1A8, 1A9, and 1A10 gene promoters:
            differential regulation through an interior-like region
  JOURNAL   J. Biol. Chem. 278 (38), 36107-36114 (2003)
   PUBMED   12847094
  REMARK    GeneRIF: UGT1A8, 1A9, and 1A10 genes are differentially regulated
```

```
            through an initiator element in their 5'-flanking regions
REFERENCE   2  (residues 1 to 530)
  AUTHORS   Elahi,A., Bendaly,J., Zheng,Z., Muscat,J.E., Richie,J.P. Jr.,
            Schantz,S.P. and Lazarus,P.
  TITLE     Detection of UGT1A10 polymorphisms and their association with
            orolaryngeal carcinoma risk
  JOURNAL   Cancer 98 (4), 872-880 (2003)
   PUBMED   12910533
  REMARK    GeneRIF: the UGT1A10 gene has several low-frequency missense
            polymorphisms and that the codon 139 polymorphism is an independent
            risk factor for orolaryngeal carcinoma in blacks
REFERENCE   3  (residues 1 to 530)
  AUTHORS   Gong,Q.H., Cho,J.W., Huang,T., Potter,C., Gholami,N., Basu,N.K.,
            Kubota,S., Carvalho,S., Pennington,M.W., Owens,I.S. and
            Popescu,N.C.
  TITLE     Thirteen UDPglucuronosyltransferase genes are encoded at the human
            UGT1 gene complex locus
  JOURNAL   Pharmacogenetics 11 (4), 357-368 (2001)
   PUBMED   11434514
REFERENCE   4  (residues 1 to 530)
  AUTHORS   Mackenzie,P.I., Owens,I.S., Burchell,B., Bock,K.W., Bairoch,A.,
            Belanger,A., Fournel-Gigleux,S., Green,M., Hum,D.W., Iyanagi,T.,
            Lancet,D., Louisot,P., Magdalou,J., Chowdhury,J.R., Ritter,J.K.,
            Schachter,H., Tephly,T.R., Tipton,K.F. and Nebert,D.W.
  TITLE     The UDP glycosyltransferase gene superfamily: recommended
            nomenclature update based on evolutionary divergence
  JOURNAL   Pharmacogenetics 7 (4), 255-269 (1997)
   PUBMED   9295054
REFERENCE   5  (residues 1 to 530)
  AUTHORS   Strassburg,C.P., Oldhafer,K., Manns,M.P. and Tukey,R.H.
  TITLE     Differential expression of the UGT1A locus in human liver, biliary,
            and gastric tissue: identification of UGT1A7 and UGT1A10
            transcripts in extrahepatic tissue
  JOURNAL   Mol. Pharmacol. 52 (2), 212-220 (1997)
   PUBMED   9271343
REFERENCE   6  (residues 1 to 530)
  AUTHORS   Jensen,L.B., Tallgren,A., Troest,T. and Jensen,S.B.
  TITLE     Effect of acupuncture on myogenic headache
  JOURNAL   Scand J Dent Res 85 (6), 456-470 (1977)
   PUBMED   271343
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from BC020971.1.
FEATURES             Location/Qualifiers
     source          1..530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
     Protein         1..530
                     /product="UDP glycosyltransferase 1 family, polypeptide
                     A10"
                     /note="UDP-glucuronosyltransferase 1A10"
     Region          26..522
                     /region_name="UDP-glucoronosyl and UDP-glucosyl
                     transferase"
                     /note="UDPGT"
                     /db_xref="CDD:22944"
     variation       139
                     /replace="K"
                     /replace="E"
                     /db_xref="dbSNP:10187694"
     variation       208
                     /replace="R"
                     /replace="W"
                     /db_xref="dbSNP:1126802"
     variation       508
                     /replace="P"
                     /replace="A"
                     /db_xref="dbSNP:1042709"
```

```
     CDS             1..530
                     /gene="UGT1A10"
                     /coded_by="NM_019075.1:70..1662"
                     /note="go_function: UDP-glucuronosyltransferase [goid
                     0003981] [evidence NR]"
                     /db_xref="GeneID:54575"
                     /db_xref="LocusID:54575"
                     /db_xref="MIM:606435"
ORIGIN
        1 maragwtspv plcvcllltc gfaeagkllv vpmdgshwft mqsvveklil rghevvvvmp
       61 evswqlersl nctvktysts ytledqnref mvfahaqwka qaqsifsllm ssssgfldlf
      121 fshcrslfnd rklveylkes sfdavfldpf dtcglivaky fslpsvvftr gifchhleeg
      181 aqcpaplsyv pndllgfsda mtfkervwnh ivhledhlfc qylfrnalei aseilqtpvt
      241 aydlyshtsi wllrtdfvld ypkpvmpnmi figginchqg kplpmefeay inasgehgiv
      301 vfslgsmvse ipekkamaia dalgkipqtv lwrytgtrps nlanntilvk wlpqndllgh
      361 pmtrafitha gshgvyesic ngvpmvmmpl fgdqmdnakr metkgagvtl nvlemtsedl
      421 enalkavind ksykenimrl sslhkdrpve pldlavfwve fvmrhkgaph lrpaahdltw
      481 yqyhsldvig fllavvltva fitfkccayg yrkclgkkgr vkkahkskth
//
```

4: NP_061966. UDP glycosyltrans...[gi:13487900]

BLink, Domains, Links

```
LOCUS       NP_061966                534 aa            linear   PRI 25-JAN-2004
DEFINITION  UDP glycosyltransferase 1 family, polypeptide A3 [Homo sapiens].
ACCESSION   NP_061966
VERSION     NP_061966.1  GI:13487900
DBSOURCE    REFSEQ: accession NM_019093.2
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 534)
  AUTHORS   Zhang,T., Haws,P. and Wu,Q.
  TITLE     Multiple Variable First Exons: A Mechanism for Cell- and
            Tissue-Specific Gene Regulation
  JOURNAL   Genome Res. 14 (1), 79-89 (2004)
   PUBMED   14672974
REFERENCE   2  (residues 1 to 534)
  AUTHORS   Mackenzie,P.I., Owens,I.S., Burchell,B., Bock,K.W., Bairoch,A.,
            Belanger,A., Fournel-Gigleux,S., Green,M., Hum,D.W., Iyanagi,T.,
            Lancet,D., Louisot,P., Magdalou,J., Chowdhury,J.R., Ritter,J.K.,
            Schachter,H., Tephly,T.R., Tipton,K.F. and Nebert,D.W.
  TITLE     The UDP glycosyltransferase gene superfamily: recommended
            nomenclature update based on evolutionary divergence
  JOURNAL   Pharmacogenetics 7 (4), 255-269 (1997)
   PUBMED   9295054
REFERENCE   3  (residues 1 to 534)
  AUTHORS   Ritter,J.K., Chen,F., Sheen,Y.Y., Tran,H.M., Kimura,S.,
            Yeatman,M.T. and Owens,I.S.
  TITLE     A novel complex locus UGT1 encodes human bilirubin, phenol, and
            other UDP-glucuronosyltransferase isozymes with identical carboxyl
            termini
  JOURNAL   J. Biol. Chem. 267 (5), 3257-3261 (1992)
   PUBMED   1339448
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from AY435138.1.
FEATURES             Location/Qualifiers
     source          1..534
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
     Protein         1..534
                     /product="UDP glycosyltransferase 1 family, polypeptide
                     A3"
     variation       11
                     /replace="R"
                     /replace="W"
```

```
                     /db_xref="dbSNP:3821242"
     Region          24..510
                     /region_name="UDP-glucuronosyl and UDP-glucosyl
                     transferase [Carbohydrate transport and metabolism, Energy
                     production and conversion]"
                     /note="KOG1192"
                     /db_xref="CDD:18981"
     Region          29..526
                     /region_name="UDP-glucoronosyl and UDP-glucosyl
                     transferase"
                     /note="UDPGT"
                     /db_xref="CDD:24386"
     variation       47
                     /replace="A"
                     /replace="V"
                     /db_xref="dbSNP:6431625"
     variation       512
                     /replace="P"
                     /replace="A"
                     /db_xref="dbSNP:1042709"
     CDS             1..534
                     /gene="UGT1A3"
                     /coded_by="NM_019093.2:1..1605"
                     /note="go_component: microsome [goid 0005792] [evidence
                     IEA];
                     go_component: integral to membrane [goid 0016021]
                     [evidence IEA];
                     go_function: UDP-glycosyltransferase activity [goid
                     0008194] [evidence TAS] [pmid 9295054];
                     go_function: glucuronosyltransferase activity [goid
                     0015020] [evidence IEA];
                     go_process: metabolism [goid 0008152] [evidence TAS] [pmid
                     9295054]"
                     /db_xref="GeneID:54659"
                     /db_xref="LocusID:54659"
                     /db_xref="MIM:606428"
ORIGIN
        1 matglqvplp wlatglllll svqpwaesgk vlvvpidgsh wlsmrevlre lharghqavv
       61 ltpevnmhik eenfftltty aiswtqdefd rhvlghtqly fetehflkkf frsmamlnnm
      121 slvyhrscve llhnealirh lnatsfdvvl tdpvnlcaav lakylsiptv fflrnipcdl
      181 dfkgtqcpnp ssyiprlltt nsdhmtfmqr vknmlyplal syichafsap yaslaselfq
      241 revsvvdils hasvwlfrgd fvmdyprpim pnmvfiggin canrkplsqe feayinasge
      301 hgivvfslgs mvseipekka maiadalgki pqtvlwrytg trpsnlannt ilvkwlpqnd
      361 llghpmtraf ithagshgvy esicngvpmv mmplfgdqmd nakrmetkga gvtlnvlemt
      421 sedlenalka vindksyken imrlsslhkd rpvepldlav fwvefvmrhk gaphlrpaah
      481 dltwyqyhsl dvigfllavv ltvafitfkc caygyrkclg kkgrvkkahk skth
//
```

5: NP_066962. UDP glycosyltrans...[gi:10863941]

BLink, Domains, Links

```
LOCUS       NP_066962                528 aa            linear   PRI 21-DEC-2003
DEFINITION  UDP glycosyltransferase 2 family, polypeptide B4;
            UDP-glucuronyltransferase, family 2, beta-4 [Homo sapiens].
ACCESSION   NP_066962
VERSION     NP_066962.1  GI:10863941
DBSOURCE    REFSEQ: accession NM_021139.1
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 528)
  AUTHORS   Barbier,O., Duran-Sandoval,D., Pineda-Torra,I., Kosykh,V.,
            Fruchart,J.C. and Staels,B.
  TITLE     Peroxisome proliferator-activated receptor alpha induces hepatic
            expression of the human bile acid glucuronidating
            UDP-glucuronosyltransferase 2B4 enzyme
  JOURNAL   J. Biol. Chem. 278 (35), 32852-32860 (2003)
   PUBMED   12810707
```

```
  REMARK     GeneRIF: UGT2B4 expression is regulated by PPARalpha
REFERENCE    2  (residues 1 to 528)
  AUTHORS    Barbier,O., Torra,I.P., Sirvent,A., Claudel,T., Blanquart,C.,
             Duran-Sandoval,D., Kuipers,F., Kosykh,V., Fruchart,J.C. and
             Staels,B.
  TITLE      FXR induces the UGT2B4 enzyme in hepatocytes: a potential mechanism
             of negative feedback control of FXR activity
  JOURNAL    Gastroenterology 124 (7), 1926-1940 (2003)
   PUBMED    12806625
  REMARK     GeneRIF: Farnesoid X receptor (FXR) induces the UGT2B4 enzyme in
             hepatocytes; this study identifies UGT2B4 as a novel FXR target
             gene.
REFERENCE    3  (residues 1 to 528)
  AUTHORS    Mackenzie,P.I., Owens,I.S., Burchell,B., Bock,K.W., Bairoch,A.,
             Belanger,A., Fournel-Gigleux,S., Green,M., Hum,D.W., Iyanagi,T.,
             Lancet,D., Louisot,P., Magdalou,J., Chowdhury,J.R., Ritter,J.K.,
             Schachter,H., Tephly,T.R., Tipton,K.F. and Nebert,D.W.
  TITLE      The UDP glycosyltransferase gene superfamily: recommended
             nomenclature update based on evolutionary divergence
  JOURNAL    Pharmacogenetics 7 (4), 255-269 (1997)
   PUBMED    9295054
REFERENCE    4  (residues 1 to 528)
  AUTHORS    Monaghan,G., Clarke,D.J., Povey,S., See,C.G., Boxer,M. and
             Burchell,B.
  TITLE      Isolation of a human YAC contig encompassing a cluster of UGT2
             genes and its regional localization to chromosome 4q13
  JOURNAL    Genomics 23 (2), 496-499 (1994)
   PUBMED    7835904
REFERENCE    5  (residues 1 to 528)
  AUTHORS    Jin,C.J., Miners,J.O., Lillywhite,K.J. and Mackenzie,P.I.
  TITLE      cDNA cloning and expression of two new members of the human liver
             UDP-glucuronosyltransferase 2B subfamily
  JOURNAL    Biochem. Biophys. Res. Commun. 194 (1), 496-503 (1993)
   PUBMED    8333863
REFERENCE    6  (residues 1 to 528)
  AUTHORS    Ritter,J.K., Chen,F., Sheen,Y.Y., Lubet,R.A. and Owens,I.S.
  TITLE      Two human liver cDNAs encode UDP-glucuronosyltransferases with 2
             log differences in activity toward parallel substrates including
             hyodeoxycholic acid and certain estrogen derivatives
  JOURNAL    Biochemistry 31 (13), 3409-3414 (1992)
   PUBMED    1554722
REFERENCE    7  (residues 1 to 528)
  AUTHORS    Jackson,M.R., McCarthy,L.R., Harding,D., Wilson,S., Coughtrie,M.W.
             and Burchell,B.
  TITLE      Cloning of a human liver microsomal UDP-glucuronosyltransferase
             cDNA
  JOURNAL    Biochem. J. 242 (2), 581-588 (1987)
   PUBMED    3109396
COMMENT      PROVISIONAL REFSEQ: This record has not yet been subject to final
             NCBI review. The reference sequence was derived from Y00317.1.
FEATURES             Location/Qualifiers
     source          1..528
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="4"
                     /map="4q13"
     Protein         1..528
                     /product="UDP glycosyltransferase 2 family, polypeptide
                     B4"
                     /EC_number="2.4.1.17"
                     /note="UDP-glucuronyltransferase, family 2, beta-4"
     sig_peptide     1..23
                     /note="putative"
     mat_peptide     24..528
                     /product="mature UDPGT (AA 1-505) (EC 2.4.1.17)"
     Region          24..526
                     /region_name="UDP-glucoronosyl and UDP-glucosyl
                     transferase"
                     /note="UDPGT"
```

```
                     /db_xref="CDD:22944"
     CDS             1..528
                     /gene="UGT2B4"
                     /coded_by="NM_021139.1:38..1624"
                     /note="go_component: microsome [goid 0005792] [evidence
                     NAS];
                     go_component: integral to membrane [goid 0016021]
                     [evidence IEA];
                     go_function: glucuronosyltransferase activity [goid
                     0015020] [evidence IEA];
                     go_process: xenobiotic metabolism [goid 0006805] [evidence
                     IDA] [pmid 8333863];
                     go_process: estrogen catabolism [goid 0006711] [evidence
                     IDA] [pmid 8333863]"
                     /db_xref="GeneID:7363"
                     /db_xref="LocusID:7363"
                     /db_xref="MIM:600067"
ORIGIN
        1 msmkwtsall liqlscyfss gscgkvlvwp tefshwmnik tildelvqrg hevtvlassa
       61 sisfdpnsps tlkfevypvs ltktefedii kqlvkrwael pkdtfwsyfs qvqeimwtfn
      121 dilrkfckdi vsnkklmkkl qesrfdvvla davfpfgell aellkipfvy rprfspgyai
      181 ekhsggllfp psyvpvvmse lsdqmtfier vknmiyvlyf efwfqifdmk kwdqfysevl
      241 grpttlsetm akadiwlirn ywdfqfphpl lpnvefvggl hckpakplpk emeefvqssg
      301 engvvvfslg smvsntseer anviasalak ipqkvlwrfd gnkpdtlgln trlykwipqn
      361 dllghpktra fithggangi ykaispripm vgvplfadqp dniahmkakg aavsldfhtm
      421 sstdllnalk tvindplyke namklsrihh dqpvkpldra vfwiefvmrh kgakhlrvaa
      481 hdltwfqyhs ldvtgfllac vatvifiitk clfcvwkfvr tgkkgkrd
//
```

6: NP_000454. UDP glycosyltrans...[gi:8850236]

BLink, Domains, Links

```
LOCUS       NP_000454                533 aa            linear   PRI 20-DEC-2003
DEFINITION  UDP glycosyltransferase 1 family, polypeptide A1 [Homo sapiens].
ACCESSION   NP_000454
VERSION     NP_000454.1  GI:8850236
DBSOURCE    REFSEQ: accession NM_000463.1
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 533)
  AUTHORS   Ohnishi,A. and Emi,Y.
  TITLE     Rapid proteasomal degradation of translocation-deficient
            UDP-glucuronosyltransferase 1A1 proteins in patients with
            Crigler-Najjar type II
  JOURNAL   Biochem. Biophys. Res. Commun. 310 (3), 735-741 (2003)
   PUBMED   14550264
  REMARK    GeneRIF: a translocation-deficient UGT1A1 protein is involved in
            Crigler-Najjar syndrome
REFERENCE   2  (residues 1 to 533)
  AUTHORS   Innocenti,F. and Ratain,M.J.
  TITLE     Irinotecan treatment in cancer patients with UGT1A1 polymorphisms
  JOURNAL   Oncology (Huntington, N.Y.) 17 (5 Suppl 5), 52-55 (2003)
   PUBMED   12800608
  REMARK    GeneRIF: genetic variation in genes involoved in the disposition of
            anticancer agents might alter patient't outcome.
REFERENCE   3  (residues 1 to 533)
  AUTHORS   Kohle,C., Mohrle,B., Munzel,P.A., Schwab,M., Wernet,D., Badary,O.A.
            and Bock,K.W.
  TITLE     Frequent co-occurrence of the TATA box mutation associated with
            Gilbert's syndrome (UGT1A1*28) with other polymorphisms of the
            UDP-glucuronosyltransferase-1 locus (UGT1A6*2 and UGT1A7*3) in
            Caucasians and Egyptians
  JOURNAL   Biochem. Pharmacol. 65 (9), 1521-1527 (2003)
   PUBMED   12732365
  REMARK    GeneRIF: Frequent haplotypes containing several UGT1 allelic
            variants should be taken into account in studies on the association
            between diseases, abnormal drug reactions, and UGT1 family
```

polymorphisms.
REFERENCE 4 (residues 1 to 533)
AUTHORS Heeney,M.M., Howard,T.A., Zimmerman,S.A. and Ware,R.E.
TITLE UGT1A promoter polymorphisms influence bilirubin response to hydroxyurea therapy in sickle cell anemia
JOURNAL J. Lab. Clin. Med. 141 (4), 279-282 (2003)
PUBMED 12677174
REMARK GeneRIF: The UGT1A promoter polymorphism is a powerful nonglobin genetic modifier in Sickle Cell Anemia that influences serum bilirubin both at baseline and on hydroxyurea therapy.
REFERENCE 5 (residues 1 to 533)
AUTHORS Yueh,M.F., Huang,Y.H., Hiller,A., Chen,S., Nguyen,N. and Tukey,R.H.
TITLE Involvement of the xenobiotic response element (XRE) in Ah receptor-mediated induction of human UDP-glucuronosyltransferase 1A1
JOURNAL J. Biol. Chem. 278 (17), 15001-15006 (2003)
PUBMED 12566446
REMARK GeneRIF: UGT1A1 induction by ligand binding to the Ah receptor was regionalized to a UGT1A1 enhancer region containing a xenobiotic response element (XRE) at -3381/-3299
REFERENCE 6 (residues 1 to 533)
AUTHORS Basu,N.K., Kole,L. and Owens,I.S.
TITLE Evidence for phosphorylation requirement for human bilirubin UDP-glucuronosyltransferase (UGT1A1) activity
JOURNAL Biochem. Biophys. Res. Commun. 303 (1), 98-104 (2003)
PUBMED 12646172
REMARK GeneRIF: human bilirubin UDP-glucuronosyltransferase requires phosphorylation for activity
REFERENCE 7 (residues 1 to 533)
AUTHORS Bosma,P.J.
TITLE Inherited disorders of bilirubin metabolism
JOURNAL J. Hepatol. 38 (1), 107-117 (2003)
PUBMED 12480568
REMARK GeneRIF: Crigler Najar syndrome and Gilbert syndrome caused by deficiency in hepatic glucuronidation of bilirubin resulting from mutation of UGTqA1 gene. (review)
REFERENCE 8 (residues 1 to 533)
AUTHORS Basten,G.P., Bao,Y. and Williamson,G.
TITLE Sulforaphane and its glutathione conjugate but not sulforaphane nitrile induce UDP-glucuronosyl transferase (UGT1A1) and glutathione transferase (GSTA1) in cultured cells
JOURNAL Carcinogenesis 23 (8), 1399-1404 (2002)
PUBMED 12151360
REMARK GeneRIF: induction in cultured tumor cells by sulforaphane and its glutathione conjugate
REFERENCE 9 (residues 1 to 533)
AUTHORS Huang,C.S., Chang,P.F., Huang,M.J., Chen,E.S. and Chen,W.C.
TITLE Glucose-6-phosphate dehydrogenase deficiency, the UDP-glucuronosyl transferase 1A1 gene, and neonatal hyperbilirubinemia
JOURNAL Gastroenterology 123 (1), 127-133 (2002)
PUBMED 12105841
REMARK GeneRIF: results indicate that carriage of the homozygous 211 G to A variation within the coding region is an additive risk factor for neonatal hyperbilirubinemia in G6PD-deficient Taiwanese male neonates.
REFERENCE 10 (residues 1 to 533)
AUTHORS Sugatani,J., Yamakawa,K., Yoshinari,K., Machida,T., Takagi,H., Mori,M., Kakizaki,S., Sueyoshi,T., Negishi,M. and Miwa,M.
TITLE Identification of a defect in the UGT1A1 gene promoter and its association with hyperbilirubinemia
JOURNAL Biochem. Biophys. Res. Commun. 292 (2), 492-497 (2002)
PUBMED 11906189
REMARK GeneRIF: polymorphism in the UGT1A1 gene promoter and its association with hyperbilirubinemia
REFERENCE 11 (residues 1 to 533)
AUTHORS Fertrin,K.Y., Goncalves,M.S., Saad,S.T. and Costa,F.F.
TITLE Frequencies of UDP-glucuronosyltransferase 1 (UGT1A1) gene promoter polymorphisms among distinct ethnic groups from Brazil
JOURNAL Am. J. Med. Genet. 108 (2), 117-119 (2002)

PUBMED 11857560
REMARK GeneRIF: The high frequencies of (TA)(7) polymorphism among the three groups confirm previous data that this polymorphism is very ancient and appears to be distributed throughout the world.
REFERENCE 12 (residues 1 to 533)
AUTHORS Sappal,B.S., Ghosh,S.S., Shneider,B., Kadakol,A., Chowdhury,J.R. and Chowdhury,N.R.
TITLE A novel intronic mutation results in the use of a cryptic splice acceptor site within the coding region of UGT1A1, causing Crigler-Najjar syndrome type 1
JOURNAL Mol. Genet. Metab. 75 (2), 134-142 (2002)
PUBMED 11855932
REMARK GeneRIF: A novel G > A mutation at the splice acceptor site in intron 4, causing Crigler-Najjar syndrome type 1
REFERENCE 13 (residues 1 to 533)
AUTHORS Passon,R.G., Howard,T.A., Zimmerman,S.A., Schultz,W.H. and Ware,R.E.
TITLE Influence of bilirubin uridine diphosphate-glucuronosyltransferase 1A promoter polymorphisms on serum bilirubin levels and cholelithiasis in children with sickle cell anemia
JOURNAL J. Pediatr. Hematol. Oncol. 23 (7), 448-451 (2001)
PUBMED 11878580
REMARK GeneRIF: Polymorphisms in UGT1A1 causes cholelithiasis and a modifier of bilirubin metabolism
REFERENCE 14 (residues 1 to 533)
AUTHORS Mackenzie,P.I., Owens,I.S., Burchell,B., Bock,K.W., Bairoch,A., Belanger,A., Fournel-Gigleux,S., Green,M., Hum,D.W., Iyanagi,T., Lancet,D., Louisot,P., Magdalou,J., Chowdhury,J.R., Ritter,J.K., Schachter,H., Tephly,T.R., Tipton,K.F. and Nebert,D.W.
TITLE The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence
JOURNAL Pharmacogenetics 7 (4), 255-269 (1997)
PUBMED 9295054
REFERENCE 15 (residues 1 to 533)
AUTHORS Mojarrabi,B., Butler,R. and Mackenzie,P.I.
TITLE cDNA cloning and characterization of the human UDP glucuronosyltransferase, UGT1A3
JOURNAL Biochem. Biophys. Res. Commun. 225 (3), 785-790 (1996)
PUBMED 8780690
REFERENCE 16 (residues 1 to 533)
AUTHORS Ritter,J.K., Chen,F., Sheen,Y.Y., Tran,H.M., Kimura,S., Yeatman,M.T. and Owens,I.S.
TITLE A novel complex locus UGT1 encodes human bilirubin, phenol, and other UDP-glucuronosyltransferase isozymes with identical carboxyl termini
JOURNAL J. Biol. Chem. 267 (5), 3257-3261 (1992)
PUBMED 1339448
REFERENCE 17 (residues 1 to 533)
AUTHORS Ritter,J.K., Crawford,J.M. and Owens,I.S.
TITLE Cloning of two human liver bilirubin UDP-glucuronosyltransferase cDNAs with expression in COS-1 cells
JOURNAL J. Biol. Chem. 266 (2), 1043-1047 (1991)
PUBMED 1898728
COMMENT PROVISIONAL REFSEQ: This record has not yet been subject to final NCBI review. The reference sequence was derived from M57899.1.
FEATURES Location/Qualifiers
source 1..533
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="2"
/map="2q37"
Protein 1..533
/product="UDP glycosyltransferase 1 family, polypeptide A1"
mat_peptide 1..533
/gene="UGT1A1"
/product="UDP glycosyltransferase 1 family, polypeptide A1"
/EC_number="2.4.1.17"

```
                     /standard_name="bilirubin UDP-glucuronosyltransferase
                     isozyme 1"
                     /note="G00-120-007"
     sig_peptide     1..533
                     /gene="UGT1A1"
                     /note="G00-120-007"
     Region          28..525
                     /region_name="UDP-glucoronosyl and UDP-glucosyl
                     transferase"
                     /note="UDPGT"
                     /db_xref="CDD:22944"
     variation       71
                     /replace="R"
                     /replace="G"
                     /db_xref="dbSNP:4148323"
     variation       511
                     /replace="P"
                     /replace="A"
                     /db_xref="dbSNP:1042709"
     CDS             1..533
                     /gene="UGT1A1"
                     /coded_by="NM_000463.1:16..1617"
                     /note="go_component: microsome [goid 0005792] [evidence
                     IEA];
                     go_component: integral to membrane [goid 0016021]
                     [evidence IEA];
                     go_function: UDP-glucuronosyltransferase [goid 0003981]
                     [evidence E];
                     go_function: glucuronosyltransferase activity [goid
                     0015020] [evidence IEA];
                     go_process: bilirubin conjugation [goid 0006789] [evidence
                     TAS] [pmid 1339448];
                     go_process: estrogen metabolism [goid 0008210] [evidence
                     TAS] [pmid 8780690];
                     go_process: digestion [goid 0007586] [evidence NR] [pmid
                     1898728];
                     go_process: metabolism [goid 0008152] [evidence IEA]"
                     /db_xref="GeneID:54658"
                     /db_xref="LocusID:54658"
                     /db_xref="MIM:191740"
ORIGIN
        1 mavesqggrp lvlglllcvl gpvvshagki llipvdgshw lsmlgaiqql qqrgheivvl
       61 apdaslyird gafytlktyp vpfqredvke sfvslghnvf endsflqrvi ktykkikkds
      121 amllsgcshl lhnkelmasl aessfdvmlt dpflpcspiv aqylslptvf flhalpcsle
      181 featqcpnpf syvprplssh sdhmtflqrv knmliafsqn flcdvvyspy atlaseflqr
      241 evtvqdllss asvwlfrsdf vkdyprpimp nmvfvgginc lhqnplsqef eayinasgeh
      301 givvfslgsm vseipekkam aiadalgkip qtvlwrytgt rpsnlannti lvkwlpqndl
      361 lghpmtrafi thagshgvye sicngvpmvm mplfgdqmdn akrmetkgag vtlnvlemts
      421 edlenalkav indksykeni mrlsslhkdr pvepldlavf wvefvmrhkg aphlrpaahd
      481 ltwyqyhsld vigfllavvl tvafitfkcc aygyrkclgk kgrvkkahks kth
//
```

7: NP_009051. UDP glycosyltrans...[gi:6005930]

BLink, Domains, Links

```
LOCUS       NP_009051                534 aa            linear   PRI 24-DEC-2003
DEFINITION  UDP glycosyltransferase 1 family, polypeptide A4 [Homo sapiens].
ACCESSION   NP_009051
VERSION     NP_009051.1  GI:6005930
DBSOURCE    REFSEQ: accession NM_007120.1
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 534)
  AUTHORS   Mackenzie,P.I., Owens,I.S., Burchell,B., Bock,K.W., Bairoch,A.,
            Belanger,A., Fournel-Gigleux,S., Green,M., Hum,D.W., Iyanagi,T.,
            Lancet,D., Louisot,P., Magdalou,J., Chowdhury,J.R., Ritter,J.K.,
            Schachter,H., Tephly,T.R., Tipton,K.F. and Nebert,D.W.
```

```
  TITLE     The UDP glycosyltransferase gene superfamily: recommended
            nomenclature update based on evolutionary divergence
  JOURNAL   Pharmacogenetics 7 (4), 255-269 (1997)
   PUBMED   9295054
REFERENCE   2  (residues 1 to 534)
  AUTHORS   Monaghan,G., Clarke,D.J., Povey,S., See,C.G., Boxer,M. and
            Burchell,B.
  TITLE     Isolation of a human YAC contig encompassing a cluster of UGT2
            genes and its regional localization to chromosome 4q13
  JOURNAL   Genomics 23 (2), 496-499 (1994)
   PUBMED   7835904
REFERENCE   3  (residues 1 to 534)
  AUTHORS   Monaghan,G., Povey,S., Burchell,B. and Boxer,M.
  TITLE     Localization of a bile acid UDP-glucuronosyltransferase gene
            (UGT2B) to chromosome 4 using the polymerase chain reaction
  JOURNAL   Genomics 13 (3), 908-909 (1992)
   PUBMED   1639428
REFERENCE   4  (residues 1 to 534)
  AUTHORS   Ritter,J.K., Chen,F., Sheen,Y.Y., Tran,H.M., Kimura,S.,
            Yeatman,M.T. and Owens,I.S.
  TITLE     A novel complex locus UGT1 encodes human bilirubin, phenol, and
            other UDP-glucuronosyltransferase isozymes with identical carboxyl
            termini
  JOURNAL   J. Biol. Chem. 267 (5), 3257-3261 (1992)
   PUBMED   1339448
REFERENCE   5  (residues 1 to 534)
  AUTHORS   Ritter,J.K., Crawford,J.M. and Owens,I.S.
  TITLE     Cloning of two human liver bilirubin UDP-glucuronosyltransferase
            cDNAs with expression in COS-1 cells
  JOURNAL   J. Biol. Chem. 266 (2), 1043-1047 (1991)
   PUBMED   1898728
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from M57951.1.
FEATURES             Location/Qualifiers
     source          1..534
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
     Protein         1..534
                     /product="UDP glycosyltransferase 1 family, polypeptide
                     A4"
     variation       11
                     /replace="R"
                     /replace="W"
                     /db_xref="dbSNP:3892221"
     sig_peptide     12..22
                     /gene="UGT2B"
                     /note="G00-127-753"
     mat_peptide     23..534
                     /gene="UGT2B"
                     /product="UDP glycosyltransferase 1 family, polypeptide
                     A4"
                     /EC_number="2.4.1.17"
                     /note="G00-127-753"
     variation       24
                     /replace="P"
                     /replace="T"
                     /db_xref="dbSNP:6755571"
     Region          29..526
                     /region_name="UDP-glucoronosyl and UDP-glucosyl
                     transferase"
                     /note="UDPGT"
                     /db_xref="CDD:22944"
     variation       48
                     /replace="L"
                     /replace="V"
                     /db_xref="dbSNP:2011425"
     variation       512
```

```
                     /replace="P"
                     /replace="A"
                     /db_xref="dbSNP:1042709"
     CDS             1..534
                     /gene="UGT1A4"
                     /coded_by="NM_007120.1:30..1634"
                     /note="go_component: microsome [goid 0005792] [evidence
                     IEA];
                     go_component: endoplasmic reticulum [goid 0005783]
                     [evidence TAS] [pmid 1898728];
                     go_component: integral to membrane [goid 0016021]
                     [evidence IEA];
                     go_function: glucuronosyltransferase activity [goid
                     0015020] [evidence IEA];
                     go_process: metabolism [goid 0008152] [evidence IEA]"
                     /db_xref="GeneID:54657"
                     /db_xref="LocusID:54657"
                     /db_xref="MIM:606429"
ORIGIN
        1 marglqvplp rlatglllll svqpwaesgk vlvvptdgsp wlsmrealre lharghqavv
       61 ltpevnmhik eekfftltay avpwtqkefd rvtlgytqgf fetehllkry srsmaimnnv
      121 slalhrccve llhnealirh lnatsfdvvl tdpvnlcgav lakylsipav ffwryipcdl
      181 dfkgtqcpnp ssyipklltt nsdhmtflqr vknmlyplal syichtfsap yaslaselfq
      241 revsvvdlvs yasvwlfrgd fvmdyprpim pnmvfiggin cangkplsqe feayinasge
      301 hgivvfslgs mvseipekka maiadalgki pqtvlwrytg trpsnlannt ilvkwlpqnd
      361 llghpmtraf ithagshgvy esicngvpmv mmplfgdqmd nakrmetkga gvtlnvlemt
      421 sedlenalka vindksyken imrlsslhkd rpvepldlav fwvefvmrhk gaphlrpaah
      481 dltwyqyhsl dvigfllavv ltvafitfkc caygyrkclg kkgrvkkahk skth
//
```

8: NP_001063. UDP glycosyltrans...[gi:4507815]

BLink, Domains, Links

```
LOCUS       NP_001063                 531 aa            linear   PRI 23-DEC-2003
DEFINITION  UDP glycosyltransferase 1 family, polypeptide A6 [Homo sapiens].
ACCESSION   NP_001063
VERSION     NP_001063.1  GI:4507815
DBSOURCE    REFSEQ: accession NM_001072.1
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 531)
  AUTHORS   Kohle,C., Mohrle,B., Munzel,P.A., Schwab,M., Wernet,D., Badary,O.A.
            and Bock,K.W.
  TITLE     Frequent co-occurrence of the TATA box mutation associated with
            Gilbert's syndrome (UGT1A1*28) with other polymorphisms of the
            UDP-glucuronosyltransferase-1 locus (UGT1A6*2 and UGT1A7*3) in
            Caucasians and Egyptians
  JOURNAL   Biochem. Pharmacol. 65 (9), 1521-1527 (2003)
   PUBMED   12732365
  REMARK    GeneRIF: Frequent haplotypes containing several UGT1 allelic
            variants should be taken into account in studies on the association
            between diseases, abnormal drug reactions, and UGT1 family
            polymorphisms.
REFERENCE   2  (residues 1 to 531)
  AUTHORS   Antonio,L., Xu,J., Little,J.M., Burchell,B., Magdalou,J. and
            Radominska-Pandya,A.
  TITLE     Glucuronidation of catechols by human hepatic, gastric, and
            intestinal microsomal UDP-glucuronosyltransferases (UGT) and
            recombinant UGT1A6, UGT1A9, and UGT2B7
  JOURNAL   Arch. Biochem. Biophys. 411 (2), 251-261 (2003)
   PUBMED   12623074
  REMARK    GeneRIF: These results demonstrate for the first time
            glucuronidation of catechols by gastric and intestinal microsomal
            UGTs and three human recombinant UGT isoforms.Recombinant human
            UGT1A6, 1A9, and 2B7 effectively catalyzed catechol glucuronidation
REFERENCE   3  (residues 1 to 531)
  AUTHORS   Peters,W.H., te Morsche,R.H. and Roelofs,H.M.
```

```
  TITLE     Combined polymorphisms in UDP-glucuronosyltransferases 1A1 and 1A6:
            implications for patients with Gilbert's syndrome
  JOURNAL   J. Hepatol. 38 (1), 3-8 (2003)
   PUBMED   12480553
  REMARK    GeneRIF: Most patients with Gilbert's syndrome may have
            abnormalities in glucuronidation of aspirin or coumarin- and
            dopamine-derivatives, due to this combination of UGT1A1*28 and
            UGT1A6*2 genotypes.
REFERENCE   4  (residues 1 to 531)
  AUTHORS   Senay,C., Jedlitschky,G., Terrier,N., Burchell,B., Magdalou,J. and
            Fournel-Gigleux,S.
  TITLE     The importance of cysteine 126 in the human liver
            UDP-glucuronosyltransferase UGT1A6
  JOURNAL   Biochim. Biophys. Acta 1597 (1), 90-96 (2002)
   PUBMED   12009407
  REMARK    GeneRIF: relevance of cysteine 126 in the glucuronidation process
REFERENCE   5  (residues 1 to 531)
  AUTHORS   Gong,Q.H., Cho,J.W., Huang,T., Potter,C., Gholami,N., Basu,N.K.,
            Kubota,S., Carvalho,S., Pennington,M.W., Owens,I.S. and
            Popescu,N.C.
  TITLE     Thirteen UDPglucuronosyltransferase genes are encoded at the human
            UGT1 gene complex locus
  JOURNAL   Pharmacogenetics 11 (4), 357-368 (2001)
   PUBMED   11434514
REFERENCE   6  (residues 1 to 531)
  AUTHORS   Mackenzie,P.I., Owens,I.S., Burchell,B., Bock,K.W., Bairoch,A.,
            Belanger,A., Fournel-Gigleux,S., Green,M., Hum,D.W., Iyanagi,T.,
            Lancet,D., Louisot,P., Magdalou,J., Chowdhury,J.R., Ritter,J.K.,
            Schachter,H., Tephly,T.R., Tipton,K.F. and Nebert,D.W.
  TITLE     The UDP glycosyltransferase gene superfamily: recommended
            nomenclature update based on evolutionary divergence
  JOURNAL   Pharmacogenetics 7 (4), 255-269 (1997)
   PUBMED   9295054
REFERENCE   7  (residues 1 to 531)
  AUTHORS   Ritter,J.K., Chen,F., Sheen,Y.Y., Tran,H.M., Kimura,S.,
            Yeatman,M.T. and Owens,I.S.
  TITLE     A novel complex locus UGT1 encodes human bilirubin, phenol, and
            other UDP-glucuronosyltransferase isozymes with identical carboxyl
            termini
  JOURNAL   J. Biol. Chem. 267 (5), 3257-3261 (1992)
   PUBMED   1339448
REFERENCE   8  (residues 1 to 531)
  AUTHORS   Harding,D., Jeremiah,S.J., Povey,S. and Burchell,B.
  TITLE     Chromosomal mapping of a human phenol UDP-glucuronosyltransferase,
            GNT1
  JOURNAL   Ann. Hum. Genet. 54 (Pt 1), 17-21 (1990)
   PUBMED   2108603
REFERENCE   9  (residues 1 to 531)
  AUTHORS   Harding,D., Fournel-Gigleux,S., Jackson,M.R. and Burchell,B.
  TITLE     Cloning and substrate specificity of a human phenol
            UDP-glucuronosyltransferase expressed in COS-7 cells
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 85 (22), 8381-8385 (1988)
   PUBMED   3141926
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from J04093.1.
FEATURES             Location/Qualifiers
     source          1..531
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
     Protein         1..531
                     /product="UDP glycosyltransferase 1 family, polypeptide
                     A6"
     Region          27..523
                     /region_name="UDP-glucoronosyl and UDP-glucosyl
                     transferase"
                     /note="UDPGT"
                     /db_xref="CDD:22944"
```

```
     CDS             1..531
                     /gene="UGT1A6"
                     /coded_by="NM_001072.1:88..1683"
                     /note="go_component: microsome [goid 0005792] [evidence
                     NAS];
                     go_component: integral to membrane [goid 0016021]
                     [evidence IEA];
                     go_function: glucuronosyltransferase activity [goid
                     0015020] [evidence IEA];
                     go_process: xenobiotic metabolism [goid 0006805] [evidence
                     IDA] [pmid 3141926]"
                     /db_xref="GeneID:54578"
                     /db_xref="LocusID:54578"
                     /db_xref="MIM:606431"
ORIGIN
        1 macllrsfqr isagvfflal wgmvvgdkll vvpqdgshwl smkdivevls drgheivvvv
       61 pevnlllkey kyytrkiypv pydqeelknr yqsfgnnhfa ersfltapqt eyrnnmivig
      121 lyfincqsll qdrdtlnffk eskfdalftd palpcgvila eylglpsvyl frgfpcsleh
      181 tfsrspdpvs yiprcytkfs dhmtfsqrva nflvnllepy lfyclfskye klasavlkrd
      241 vdiitlsevs vwllrydfvl eyprpvmpnm vfigginckk rkdlsqefea yinasgehgi
      301 vvfslgsmvs eipekkamai adalgknpqt vlwrytgtrp snlanntilv kwlpqndllg
      361 hpmtrafith agshgvyesi cngvpmvmmp lfgdqmdnak rmetkgagvt lnvlemtsed
      421 lenalkavin dksykenimr lsslhkdrpv epldlavfwv efvmrhkgap hlrpaahdlt
      481 wyqyhsldvi gfllavvltv afitfkccpy gypkclgkkg rvkkahkskt h
//
```

9: NP_061950. UDP glycosyltrans...[gi:41282213]

BLink, Links

```
LOCUS       NP_061950                 530 aa            linear   PRI 25-JAN-2004
DEFINITION  UDP glycosyltransferase 1 family, polypeptide A7;
            UDP-glucuronosyltransferase 1A7 [Homo sapiens].
ACCESSION   NP_061950
VERSION     NP_061950.2  GI:41282213
DBSOURCE    REFSEQ: accession NM_019077.2
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (residues 1 to 530)
  AUTHORS   Zhang,T., Haws,P. and Wu,Q.
  TITLE     Multiple Variable First Exons: A Mechanism for Cell- and
            Tissue-Specific Gene Regulation
  JOURNAL   Genome Res. 14 (1), 79-89 (2004)
   PUBMED   14672974
REFERENCE   2  (residues 1 to 530)
  AUTHORS   Kohle,C., Mohrle,B., Munzel,P.A., Schwab,M., Wernet,D., Badary,O.A.
            and Bock,K.W.
  TITLE     Frequent co-occurrence of the TATA box mutation associated with
            Gilbert's syndrome (UGT1A1*28) with other polymorphisms of the
            UDP-glucuronosyltransferase-1 locus (UGT1A6*2 and UGT1A7*3) in
            Caucasians and Egyptians
  JOURNAL   Biochem. Pharmacol. 65 (9), 1521-1527 (2003)
   PUBMED   12732365
  REMARK    GeneRIF: Frequent haplotypes containing several UGT1 allelic
            variants should be taken into account in studies on the association
            between diseases, abnormal drug reactions, and UGT1 family
            polymorphisms.
REFERENCE   3  (residues 1 to 530)
  AUTHORS   Gong,Q.H., Cho,J.W., Huang,T., Potter,C., Gholami,N., Basu,N.K.,
            Kubota,S., Carvalho,S., Pennington,M.W., Owens,I.S. and
            Popescu,N.C.
  TITLE     Thirteen UDPglucuronosyltransferase genes are encoded at the human
            UGT1 gene complex locus
  JOURNAL   Pharmacogenetics 11 (4), 357-368 (2001)
   PUBMED   11434514
REFERENCE   4  (residues 1 to 530)
  AUTHORS   Mackenzie,P.I., Owens,I.S., Burchell,B., Bock,K.W., Bairoch,A.,
            Belanger,A., Fournel-Gigleux,S., Green,M., Hum,D.W., Iyanagi,T.,
```

```
            Lancet,D., Louisot,P., Magdalou,J., Chowdhury,J.R., Ritter,J.K.,
            Schachter,H., Tephly,T.R., Tipton,K.F. and Nebert,D.W.
  TITLE     The UDP glycosyltransferase gene superfamily: recommended
            nomenclature update based on evolutionary divergence
  JOURNAL   Pharmacogenetics 7 (4), 255-269 (1997)
   PUBMED   9295054
REFERENCE   5  (residues 1 to 530)
  AUTHORS   Strassburg,C.P., Oldhafer,K., Manns,M.P. and Tukey,R.H.
  TITLE     Differential expression of the UGT1A locus in human liver, biliary,
            and gastric tissue: identification of UGT1A7 and UGT1A10
            transcripts in extrahepatic tissue
  JOURNAL   Mol. Pharmacol. 52 (2), 212-220 (1997)
   PUBMED   9271343
COMMENT     PROVISIONAL REFSEQ: This record has not yet been subject to final
            NCBI review. The reference sequence was derived from AY435142.1.
            On Jan 25, 2004 this sequence version replaced gi:29789084.
FEATURES             Location/Qualifiers
     source          1..530
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="2"
                     /map="2q37"
     Protein         1..530
                     /product="UDP glycosyltransferase 1 family, polypeptide
                     A7"
                     /note="UDP-glucuronosyltransferase 1A7"
     Region          22..506
                     /region_name="UDP-glucuronosyl and UDP-glucosyl
                     transferase [Carbohydrate transport and metabolism, Energy
                     production and conversion]"
                     /note="KOG1192"
                     /db_xref="CDD:18981"
     Region          26..522
                     /region_name="UDP-glucoronosyl and UDP-glucosyl
                     transferase"
                     /note="UDPGT"
                     /db_xref="CDD:24386"
     CDS             1..530
                     /gene="UGT1A7"
                     /coded_by="NM_019077.2:1..1593"
                     /note="go_component: extracellular [goid 0005576]
                     [evidence IEA];
                     go_function: UDP-glucuronosyltransferase [goid 0003981]
                     [evidence NR];
                     go_function: calcium ion binding [goid 0005509] [evidence
                     IEA];
                     go_function: transferase activity [goid 0016740] [evidence
                     IEA];
                     go_function: transferase activity, transferring hexosyl
                     groups [goid 0016758] [evidence IEA];
                     go_process: metabolism [goid 0008152] [evidence IEA]"
                     /db_xref="GeneID:54577"
                     /db_xref="LocusID:54577"
                     /db_xref="MIM:606432"
ORIGIN
        1 maragwtgll plyvcllltc gfakagkllv vpmdgshwft mqsvveklil rghevvvvmp
       61 evswqlgrsl nctvktysts ytledqdref mvfadarwta plrsafsllt sssngifdlf
      121 fsncrslfnd rklveylkes cfdavfldpf dacglivaky fslpsvvfar gifchyleeg
      181 aqcpaplsyv prlllgfsda mtfkervwnh imhleehlfc pyffknvlei aseilqtpvt
      241 aydlyshtsi wllrtdfvle ypkpvmpnmi figginchqg kpvpmefeay inasgehgiv
      301 vfslgsmvse ipekkamaia dalgkipqtv lwrytgtrps nlanntilvk wlpqndllgh
      361 pmtrafitha gshgvyesic ngvpmvmmpl fgdqmdnakr metkgagvtl nvlemtsedl
      421 enalkavind ksykenimrl sslhkdrpve pldlavfwve fvmrhkgaph lrpaahdltw
      481 yqyhsldvig fllavvltva fitfkccayg yrkclgkkgr vkkahkskth
//
```

10: P22310. UDP-glucuronosylt...[gi:136731] BLink, Domains, Links

```
LOCUS       P22310                    534 aa            linear   PRI 15-SEP-2003
```

DEFINITION UDP-glucuronosyltransferase 1-4 precursor, microsomal (UDP-glucuronosyltransferase 1A4) (UDPGT) (UGT1*4) (UGT1-04) (UGT1.4) (UGT-1D) (UGT1D) (Bilirubin specific UDPGT isozyme 2) (HUG-BR2).
ACCESSION P22310
VERSION P22310 GI:136731
DBSOURCE swissprot: locus UD14_HUMAN, accession P22310;
class: standard.
created: Aug 1, 1991.
sequence updated: Aug 1, 1991.
annotation updated: Sep 15, 2003.
xrefs: gi: 340136, gi: 340137, gi: 340129, gi: 459838, gi: 340127, gi: 340128, gi: 184474, gi: 184475, gi: 11118740, gi: 11118747
xrefs (non-sequence databases): GenewHGNC:12536, MIM 606429, MIM 191740, MIM 143500, MIM 218800, GO0005783, InterProIPR002213, PfamPF00201, PROSITEPS00375
KEYWORDS Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal; Multigene family; Microsome; Alternative splicing; Disease mutation.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (residues 1 to 534)
AUTHORS Ritter,J.K., Crawford,J.M. and Owens,I.S.
TITLE Cloning of two human liver bilirubin UDP-glucuronosyltransferase cDNAs with expression in COS-1 cells
JOURNAL J. Biol. Chem. 266 (2), 1043-1047 (1991)
MEDLINE 91093210
PUBMED 1898728
REMARK SEQUENCE FROM N.A.
TISSUE=Liver
REFERENCE 2 (residues 1 to 534)
AUTHORS Ritter,J.K., Chen,F., Sheen,Y.Y., Tran,H.M., Kimura,S., Yeatman,M.T. and Owens,I.S.
TITLE A novel complex locus UGT1 encodes human bilirubin, phenol, and other UDP-glucuronosyltransferase isozymes with identical carboxyl termini
JOURNAL J. Biol. Chem. 267 (5), 3257-3261 (1992)
MEDLINE 92147680
PUBMED 1339448
REMARK SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
REFERENCE 3 (residues 1 to 534)
AUTHORS Gong,Q.H., Cho,J.W., Huang,T., Potter,C., Gholami,N., Basu,N.K., Kubota,S., Carvalho,S., Pennington,M.W., Owens,I.S. and Popescu,N.C.
TITLE Thirteen UDPglucuronosyltransferase genes are encoded at the human UGT1 gene complex locus
JOURNAL Pharmacogenetics 11 (4), 357-368 (2001)
MEDLINE 21327373
PUBMED 11434514
REMARK SEQUENCE FROM N.A.
REFERENCE 4 (residues 1 to 534)
AUTHORS Bosma,P.J., Chowdhury,J.R., Huang,T.J., Lahiri,P., Elferink,R.P., Van Es,H.H., Lederstein,M., Whitington,P.F., Jansen,P.L. and Chowdhury,N.R.
TITLE Mechanisms of inherited deficiencies of multiple UDP-glucuronosyltransferase isoforms in two patients with Crigler-Najjar syndrome, type I
JOURNAL FASEB J. 6 (10), 2859-2863 (1992)
MEDLINE 92339803
PUBMED 1634050
REMARK VARIANT CN-I PHE-376.
REFERENCE 5 (residues 1 to 534)
AUTHORS Aono,S., Yamada,Y., Keino,H., Hanada,N., Nakagawa,T., Sasaoka,Y., Yazawa,T., Sato,H. and Koiwai,O.
TITLE Identification of defect in the genes for bilirubin UDP-glucuronosyl-transferase in a patient with Crigler-Najjar syndrome type II

```
  JOURNAL   Biochem. Biophys. Res. Commun. 197 (3), 1239-1244 (1993)
  MEDLINE   94107323
   PUBMED   8280139
  REMARK    VARIANTS CN-II PRO-132 AND ASP-487.
REFERENCE   6  (residues 1 to 534)
  AUTHORS   Moghrabi,N., Clarke,D.J., Boxer,M. and Burchell,B.
  TITLE     Identification of an A-to-G missense mutation in exon 2 of the UGT1
            gene complex that causes Crigler-Najjar syndrome type 2
  JOURNAL   Genomics 18 (1), 171-173 (1993)
  MEDLINE   94102756
   PUBMED   8276413
  REMARK    VARIANT CN-II ARG-332.
COMMENT     ---------------------------------------------------------------------
            This SWISS-PROT entry is copyright. It is produced through a
            collaboration between the Swiss Institute of Bioinformatics and
            the EMBL outstation - the European Bioinformatics Institute.
            The original entry is available from http://www.expasy.ch/sprot
            and http://www.ebi.ac.uk/sprot
            ---------------------------------------------------------------------.
            [FUNCTION] UDPGT is of major importance in the conjugation and
            subsequent elimination of potentially toxic xenobiotics and
            endogenous compounds. This isoform glucuronidates bilirubin
            IX-alpha to form both the IX-alpha-C8 and IX-alpha-C12
            monoconjugates and diconjugate.
            [CATALYTIC ACTIVITY] UDP-glucuronate + acceptor = UDP + acceptor
            beta-D-glucuronoside.
            [SUBCELLULAR LOCATION] Microsomal.
            [ALTERNATIVE PRODUCTS] Event=Alternative splicing; Named
            isoforms=1; Comment=A number of isoforms are produced. The
            different isozymes have a different N-terminal domain and a common
            C-terminal domain of 245 residues; Name=1; IsoId=P22310-1;
            Sequence=Displayed.
            [TISSUE SPECIFICITY] Expressed in liver. Not expressed in skin or
            kidney.
            [INDUCTION] By phenobarbital.
            [DISEASE] THE GILBERT'S SYNDROME IS SHOWN TO OCCUR AS A CONSEQUENCE
            OF REDUCED BILIRUBIN TRANSFERASE ACTIVITY. THE DISORDER, IS MOST
            OFTEN DETECTED IN YOUNG ADULTS WITH VAGUE NONSPECIFIC COMPLAINTS. A
            MORE SEVERE INHERITABLE DEFICIENCY IN BILIRUBIN ACTIVITY EXIST IN
            CRIGLER-NAJJAR (CN): PATIENTS WITH TYPE I (RECESSIVE TRAIT) HAVE
            SEVERE HYPERBILIRUBINEMIA AND USUALLY DIE OF KERNICTERUS (BILIRUBIN
            ACCUMULATION IN THE BASAL GANGLIA AND BRAINSTEM NUCLEI) WITHIN THE
            FIRST YEAR OF LIFE. PATIENTS WITH TYPE II (DOMINANT TRAIT) HAVE
            LESS SEVERE HYPERBILIRUBINEMIA AND USUALLY SURVIVE INTO ADULTHOOD
            WITHOUT NEUROLOGIC DAMAGE. PHENOBARBITAL, WHICH INDUCES THE
            PARTIALLY DEFICIENT GLUCURONYL TRANSFERASE, CAN DIMINISH THE
            JAUNDICE.
            [SIMILARITY] Belongs to the UDP-glycosyltransferase family.
FEATURES                Location/Qualifiers
     source             1..534
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
     gene               1..534
                        /gene="UGT1A4"
                        /note="synonyms: UGT1, GNT1"
     Protein            1..534
                        /gene="UGT1A4"
                        /product="UDP-glucuronosyltransferase 1-4 precursor,
                        microsomal"
                        /EC_number="2.4.1.17"
     Region             1..28
                        /gene="UGT1A4"
                        /region_name="Signal"
                        /note="POTENTIAL."
     Region             29..534
                        /gene="UGT1A4"
                        /region_name="Mature chain"
                        /note="UDP-GLUCURONOSYLTRANSFERASE 1-4."
     Site               119
```

```
                     /gene="UGT1A4"
                     /site_type="glycosylation"
                     /note="N-LINKED (GLCNAC...) (POTENTIAL)."
     Region          132
                     /gene="UGT1A4"
                     /region_name="Variant"
                     /note="L -> P (IN CRIGLER-NAJJAR TYPE II).
                     /FTId=VAR_009506."
     Site            142
                     /gene="UGT1A4"
                     /site_type="glycosylation"
                     /note="N-LINKED (GLCNAC...) (POTENTIAL)."
     Site            296
                     /gene="UGT1A4"
                     /site_type="glycosylation"
                     /note="N-LINKED (GLCNAC...) (POTENTIAL)."
     Region          332
                     /gene="UGT1A4"
                     /region_name="Variant"
                     /note="Q -> R (IN CRIGLER-NAJJAR TYPE II).
                     /FTId=VAR_007710."
     Site            348
                     /gene="UGT1A4"
                     /site_type="glycosylation"
                     /note="N-LINKED (GLCNAC...) (POTENTIAL)."
     Region          376
                     /gene="UGT1A4"
                     /region_name="Variant"
                     /note="S -> F (IN CRIGLER-NAJJAR TYPE I).
                     /FTId=VAR_007711."
     Region          487
                     /gene="UGT1A4"
                     /region_name="Variant"
                     /note="Y -> D (IN CRIGLER-NAJJAR TYPE II).
                     /FTId=VAR_009507."
     Region          492..508
                     /gene="UGT1A4"
                     /region_name="Transmembrane region"
                     /note="POTENTIAL."
ORIGIN
        1 marglqvplp rlatglllll svqpwaesgk vlvvptdgsp wlsmrealre lharghqavv
       61 ltpevnmhik eekfftltay avpwtqkefd rvtlgytqgf fetehllkry srsmaimnnv
      121 slalhrccve llhnealirh lnatsfdvvl tdpvnlcgav lakylsipav ffwryipcdl
      181 dfkgtqcpnp ssyipklltt nsdhmtflqr vknmlyplal syichtfsap yaslaselfq
      241 revsvvdlvs yasvwlfrgd fvmdyprpim pnmvfiggin cangkplsqe feayinasge
      301 hgivvfslgs mvseipekka maiadalgki pqtvlwrytg trpsnlannt ilvkwlpqnd
      361 llghpmtraf ithagshgvy esicngvpmv mmplfgdqmd nakrmetkga gvtlnvlemt
      421 sedlenalka vindksyken imrlsslhkd rpvepldlav fwvefvmrhk gaphlrpaah
      481 dltwyqyhsl dvigfllavv ltvafitfkc caygyrkclg kkgrvkkahk skth
//
```

Display | GenPept | Show: 20 | Send to | File | Features

Items 1-10 of 10

One page.

Jan 29 2004 15:38:25